

Article

Not peer-reviewed version

AI-Enhanced Urban Building Energy Modeling for Health-Driven Decarbonization in Vulnerable Communities

[Narjes Abbasabadi](#)*, [Teresa F. Moroseos](#), [Mehdi Ashayeri](#), Christopher Meek

Posted Date: 22 April 2026

doi: 10.20944/preprints202604.1576.v1

Keywords: urban building energy modeling; machine learning; health-driven energy retrofits; indoor environmental quality; decarbonization



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

AI-Enhanced Urban Building Energy Modeling for Health-Driven Decarbonization in Vulnerable Communities

Narjes Abbasabadi ^{1,*}, Teresa F. Moroseos ¹, Mehdi Ashayeri ² and Christopher Meek ¹

¹ Department of Architecture at the University of Washington, Seattle, WA, USA

² School of Architecture at Southern Illinois University, Carbondale, IL, USA

* Correspondence: nabbasab@uw.edu

Abstract

Retrofitting existing residential buildings is a critical strategy for achieving urban decarbonization while addressing public health disparities, particularly in communities disproportionately affected by environmental and socioeconomic stressors. This study presents a scalable urban building energy modeling framework that integrates physics-based simulations with machine learning to evaluate and prioritize health-driven retrofit strategies across residential building stocks. Synthetic datasets were generated through parametric simulations of representative building archetypes and retrofit scenarios, capturing variations in envelope performance, HVAC systems, infiltration rates, and ventilation strategies. Machine learning models were trained as surrogate predictors of building energy performance, enabling rapid evaluation of retrofit impacts. A range of algorithms—including decision trees, random decision forests, gradient boosting machines, support vector machines, k-nearest neighbors, and artificial neural networks—were evaluated. An artificial neural network implemented as a multilayer perceptron was selected for further analysis due to its strong predictive performance ($R^2 = 0.94$) and ability to capture complex nonlinear relationships among retrofit variables. The final model used the Port optimization algorithm for stable convergence and improved generalization. The framework is applied to Seattle's Duwamish Valley, a community experiencing disproportionate environmental and health burdens. The results highlight retrofit priorities—particularly infiltration reduction, HVAC upgrades, and improved envelope performance—that deliver co-benefits for energy efficiency, indoor environmental quality, and occupant health. The results demonstrate that machine learning-enhanced physics-based UBEM can significantly accelerate retrofit evaluation while preserving the interpretability of simulation-based approaches. The proposed framework provides a scalable approach for identifying health-informed retrofit pathways that support equitable urban decarbonization.

Keywords: urban building energy modeling; machine learning; health-driven energy retrofits; indoor environmental quality; decarbonization

1. Introduction

Cities worldwide face the intertwined challenges of increasing energy demand, accelerating climate change, and persistent socioeconomic and environmental inequalities [1]. These challenges are particularly pronounced in historically marginalized urban neighborhoods, where communities are disproportionately exposed to environmental hazards that adversely affect health and well-being. For instance, residents living near industrial zones—where air pollution levels are elevated—are more likely to be people of color and to experience higher rates of respiratory illness and energy burden [2]. At the same time, low-income households often spend a disproportionately large share of their income on energy despite consuming less energy on average [3]. These intersecting burdens

highlight the need for integrated solutions that simultaneously address energy efficiency, environmental exposure, and public health outcomes.

Residential retrofits represent a critical leverage point for improving health outcomes while advancing building decarbonization goals. Targeted interventions—such as improved insulation, enhanced ventilation systems, and the replacement of fossil-fuel heating systems with high-efficiency heat pumps—can reduce energy consumption while improving indoor air quality and thermal comfort [4]. However, many municipalities, housing providers, and community organizations lack the analytical tools to systematically evaluate the energy, health, and long-term performance impacts of alternative retrofit strategies. To support more equitable and evidence-based decision-making, this study develops an integrated modeling framework that combines machine learning (ML) with urban building energy modeling (UBEM) to evaluate residential retrofit strategies that simultaneously improve energy performance and occupant health. The framework is applied to Seattle’s Duwamish Valley—an environmentally burdened community characterized by aging housing stock, environmental exposures, and elevated energy burdens—to identify retrofit measures that provide the greatest combined energy and health benefits.

1.1. Energy Retrofit Strategies and Their Impacts on Occupant Health

A growing body of research demonstrates that residential retrofits can deliver both energy efficiency improvements and measurable health benefits. By improving building envelope performance, ventilation, and mechanical systems, retrofit interventions can significantly enhance indoor environmental quality (IEQ) and occupant well-being.

Measures such as air sealing and improved ventilation not only reduce building energy demand but also help limit occupant exposure to pollutants, allergens, and excess moisture that contribute to respiratory illness. Envelope improvements—including sealing gaps around doors, windows, and other penetrations—reduce uncontrolled infiltration and improve thermal stability. A U.S. Department of Energy (DOE) report found that envelope improvements can reduce residential energy consumption by approximately 6–10% in mixed climates with a payback period of 1.7–2.2 years [5]. In addition to energy benefits, reducing infiltration has been shown to limit the indoor accumulation of outdoor pollutants, particularly particulate matters (PMs) [6,7]. Elevated indoor pollutant concentrations are frequently observed in low-income housing and are associated with adverse health outcomes including asthma, impaired lung function, and increased respiratory-related mortality [8–11]. For example, a study of indoor air quality in low-income housing in Colorado found that residents in homes with high infiltration rates were more likely to have chronic cough, asthma, and asthma-like symptoms [12].

While infiltration reduction is a key component of building weatherization, it must be accompanied by adequate filtered ventilation to avoid trapping indoor pollutants and moisture [13]. Indoor contaminants can originate from multiple sources, including gas cooking, combustion heating, smoking, biological particles, and household aerosols [14]. Effective ventilation systems dilute indoor pollutants and remove excess moisture that can lead to mold growth [15]. Energy recovery ventilation (ERV) systems have emerged as an effective retrofit solution because they provide filtered outdoor air while recovering heat and moisture from exhaust air streams. ERV systems equipped with MERV-13 filters can remove common outdoor pollutants while improving ventilation efficiency. Studies of low-income housing retrofits have shown that improved building envelopes combined with ERV systems can substantially reduce asthma-related health outcomes [11].

Energy retrofits can also significantly improve thermal comfort in low-income housing by addressing issues such as poor insulation, air leaks, and inefficient heating and cooling systems. Many older homes struggle to maintain stable indoor temperatures, leading to discomfort and increased impacts to quality of life and health [16]. Retrofit interventions—including envelope improvements, air sealing, and HVAC upgrades—can stabilize indoor temperatures and improve occupant well-being. For example, a study of low-income senior housing in Phoenix, Arizona found that improved roof insulation, enhanced weatherization, and upgraded HVAC systems significantly

reduced indoor temperature variability and improved residents' reported quality of life and sleep quality [17]. Heat pump systems provide an additional opportunity to improve thermal comfort while reducing energy use. Compared with electric resistance heating or gas furnaces, heat pumps transfer heat rather than generate it, resulting in significantly higher energy efficiency. In addition to improved heating efficiency, heat pumps provide mechanical cooling, which is increasingly important as heat waves become more frequent and severe.

Reducing exposure to combustion-related indoor pollutants is another important benefit of residential retrofits. Fossil-fuel combustion appliances—including gas furnaces and water heaters—can release pollutants such as carbon monoxide and nitrogen oxides into indoor environments [13]. Exposure to these pollutants is associated with respiratory and cardiovascular health risks [18]. Transitioning to electric heat pump systems can eliminate indoor combustion sources while reducing energy demand. When combined with improved ventilation strategies, these measures can significantly improve indoor air quality and reduce health risks for vulnerable populations.

In this study, *health-driven energy efficiency* refers to retrofit strategies that reduce building energy demand while simultaneously improving indoor environmental quality factors linked to occupant health, including thermal stability, ventilation performance, and reduced infiltration of outdoor pollutants.

1.2. Urban Building Energy Modeling for Scalable Retrofit Analysis

Despite increasing recognition of the health benefits of building retrofits, integrated methods that simultaneously evaluate energy performance and health outcomes remain limited. This gap restricts the ability of decision-makers to prioritize retrofit strategies that support both climate mitigation and public health. UBEM has emerged as a powerful tool for evaluating building energy performance at neighborhood and city scales [19]. UBEM frameworks represent large building stocks using archetypes or representative building typologies and integrate heterogeneous datasets—such as building geometry, construction characteristics, occupancy assumptions, and climate data—to simulate urban-scale energy consumption [20]. These models enable scenario-based analyses that can facilitate the assessment of design, retrofit, and policy interventions at urban scales by identifying cost-effective, high-impact strategies [19,21].

Physics-based modeling forms the foundation of most UBEM approaches, simulating building energy dynamics using thermodynamic and heat transfer principles [22–25]. However, conventional physics-based UBEM models rely on fixed input parameters—including building geometry, material properties, HVAC specifications, and local climate data—and deterministic assumptions that may not capture the stochastic variability of real-world systems. Factors such as occupant behavior [20,26,27], socioeconomic factors [28–31], and microclimatic variations [32] introduce uncertainties that can lead to discrepancies between simulated and observed energy performance. In addition, physics-based UBEM simulations can be computationally intensive when applied to large building stocks, often requiring simplified archetypes that fail to capture the diversity of urban building characteristics [23,33–35].

Recent advancements in artificial intelligence (AI) and machine learning (ML) offer opportunities to address these limitations [19,36,37]. When coupled with physics-based simulations [38], ML techniques can improve predictive accuracy and substantially reduce computational requirements for urban scale retrofit scenario analysis. While physics-based models remain essential for simulating fundamental processes such as heat transfer, insulation performance, and ventilation dynamics, ML approaches can identify complex, nonlinear relationships across high-dimensional input spaces—patterns that are often missed or oversimplified in conventional models. Supervised learning approaches have been widely applied to predict energy consumption and identify patterns in energy usage across urban building stocks [39]. Various algorithms—including linear regression, artificial neural networks (ANNs), support vector machines, and random forest models—have been evaluated in UBEM contexts with varying performance depending on data characteristics and modeling objectives [34,40–44].

Deep learning methods have been used for processing complex and high-dimensional data in UBEM. Deep neural networks (DNNs), such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied for tasks including urban energy prediction, image-based audits, and time series forecasting [27,45,46]. Other studies have enhanced traditional models—for example, improving support vector machine (SVM) performance using genetic algorithms for energy prediction in smart communities [47]. Some studies show that while ANN and SVM can capture complex nonlinear relationships, they do not always outperform simpler regression approaches [48,49]. Ensemble methods such as random decision forests (RDF) have also shown strong performance in forecasting building energy demand using large behavioral datasets [50]. These findings underscore the growing utility of supervised learning in UBEM, particularly when tailored to specific building typologies, data contexts, and urban energy objectives.

ML-based surrogate modeling has emerged as an approach to overcome key limitations of traditional physics-based UBEM, particularly in the context of computational intensity, scalability, and data requirements [39]. Surrogate models approximate the behavior of high-fidelity simulation models using computationally efficient predictive algorithms trained on simulation data [51,52]. These models can significantly reduce computational costs while maintaining high predictive accuracy, enabling large-scale scenario exploration and optimization. Previous studies demonstrate substantial computational gains, including large-scale simulation acceleration using DNNs across thousands of buildings [53] and ANN-based models reducing retrofit simulation time from minutes to microseconds [54]. Prior studies demonstrate ANN-based optimization for retrofit strategy identification [55], clustering and statistical surrogate approaches for large building portfolio calibration [56,57], and ML-based meta-models across diverse climatic contexts [52]. Integrated UBEM–ML workflows have also been shown to automate simulation pipelines and substantially improve computational efficiency [58], highlighting the scalability of surrogate models for large-scale urban retrofit analysis.

This study proposes an AI-enhanced physics-based modeling framework to identify residential retrofit strategies that simultaneously reduce energy demand and improve indoor environmental conditions linked to occupant health. By integrating building physics simulations with ML surrogate modeling, the framework enables large-scale evaluation of retrofit measures that support both urban decarbonization and health-resilient housing in disadvantaged communities. This approach enables scalable analysis of retrofit strategies while preserving the physical interpretability of simulation-based methods. Applied to Seattle’s Duwamish Valley, the framework identifies retrofit interventions that simultaneously reduce energy consumption and improve indoor environmental conditions, supporting more equitable pathways toward urban decarbonization, with a particular emphasis on vulnerable communities disproportionately affected by environmental and health disparities. In this context, energy retrofits can be reframed not only as climate mitigation measures but also as public-health interventions within the built environment.

This study makes three primary contributions:

1. A hybrid physics-based and machine-learning UBEM framework that combines parametric energy simulation with surrogate modeling to enable scalable evaluation of residential retrofit strategies.
2. An interpretable machine-learning modeling approach that identifies key drivers of building energy consumption through feature importance analysis and partial dependence interpretation.
3. A health-driven retrofit prioritization perspective that connects energy efficiency improvements with indoor environmental quality and public-health considerations in disadvantaged urban communities.

2. Methodology

This study develops a hybrid UBEM framework that integrates physics-based simulations with ML to evaluate health-driven residential retrofit strategies. The methodological framework consists

of two complementary modeling layers: Layer 1 – Physics-based modeling, which generates simulation data through archetype-based parametric energy simulations; and Layer 2 – Data-driven modeling, which uses machine learning algorithms trained on the simulation outputs to predict building energy performance and interpret retrofit impacts. The data-driven models serve as machine-learning surrogate models trained on physics-based simulation outputs that approximate the input–output relationships of the physics-based simulations, enabling rapid evaluation of retrofit scenarios without repeatedly executing computationally intensive simulations. Model interpretation techniques—including feature importance analysis and partial dependence plots (PDPs)—are applied to identify key drivers of energy consumption and inform health-driven retrofit strategies. Figure 1 illustrates the overall methodological workflow.

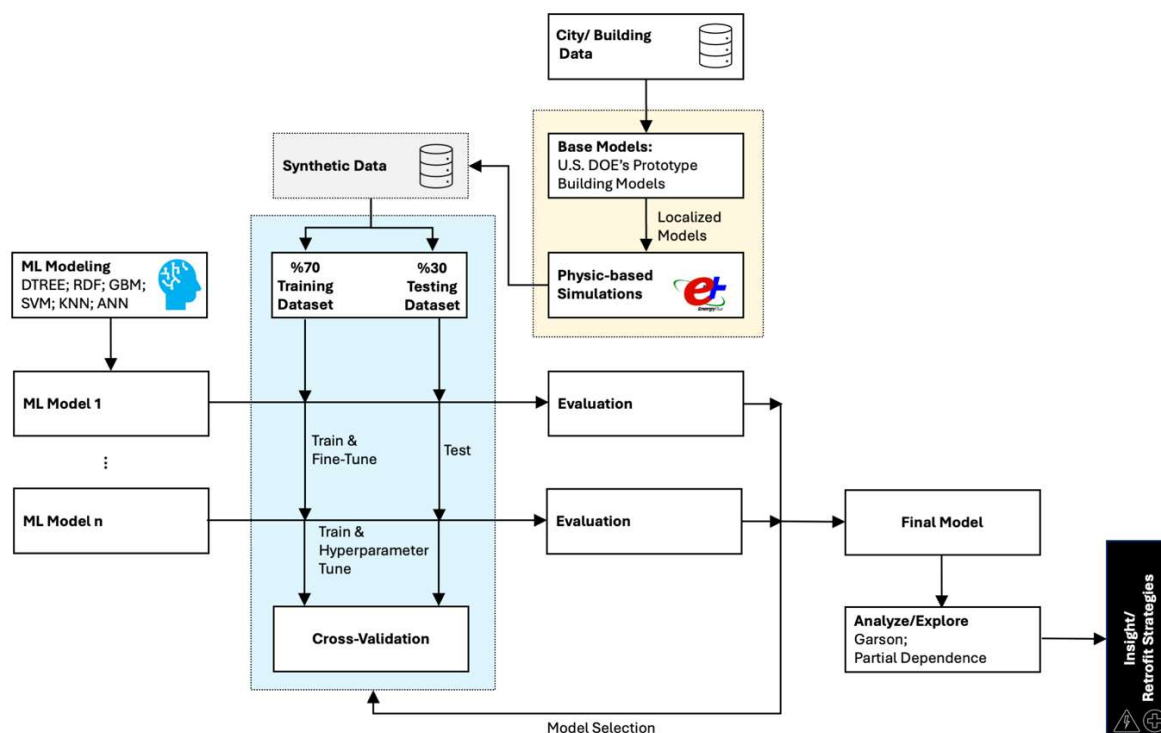


Figure 1. Workflow of the proposed research method.

2.1. Study Area and Building Stock Characterization

2.1.1. Study Area

Seattle's Duwamish Valley, Washington, was selected as the case study because it faces persistent environmental, health, and social inequities that make it highly relevant for health-driven retrofit analysis. Historically shaped by industrial land uses, the area experiences elevated pollution burdens, degraded environmental conditions, and pronounced social vulnerability [59,60]. It contains the highest concentration of contaminated sites in Seattle and experiences some of the city's poorest air quality [59]. Life expectancy in the Duwamish Valley is 73.3 years, approximately eight years below the city average, underscoring the severity of local health disparities [59].

The study focuses on South Park, one of the four neighborhoods comprising the Duwamish Valley—South Park, Georgetown, Riverton-Boulevard Park, and Tukwila (Figure 2). South Park was selected because it contains a concentration of older residential buildings, many of which are occupied by low-income households and lack modern ventilation, cooling, and high-performance envelope systems. Much of Seattle's older affordable housing stock predates the widespread adoption of whole-house mechanical ventilation, filtered outdoor air supply, and heat pump-based

heating and cooling systems. As a result, many buildings remain vulnerable to poor indoor air quality, overheating, and inefficient energy performance.

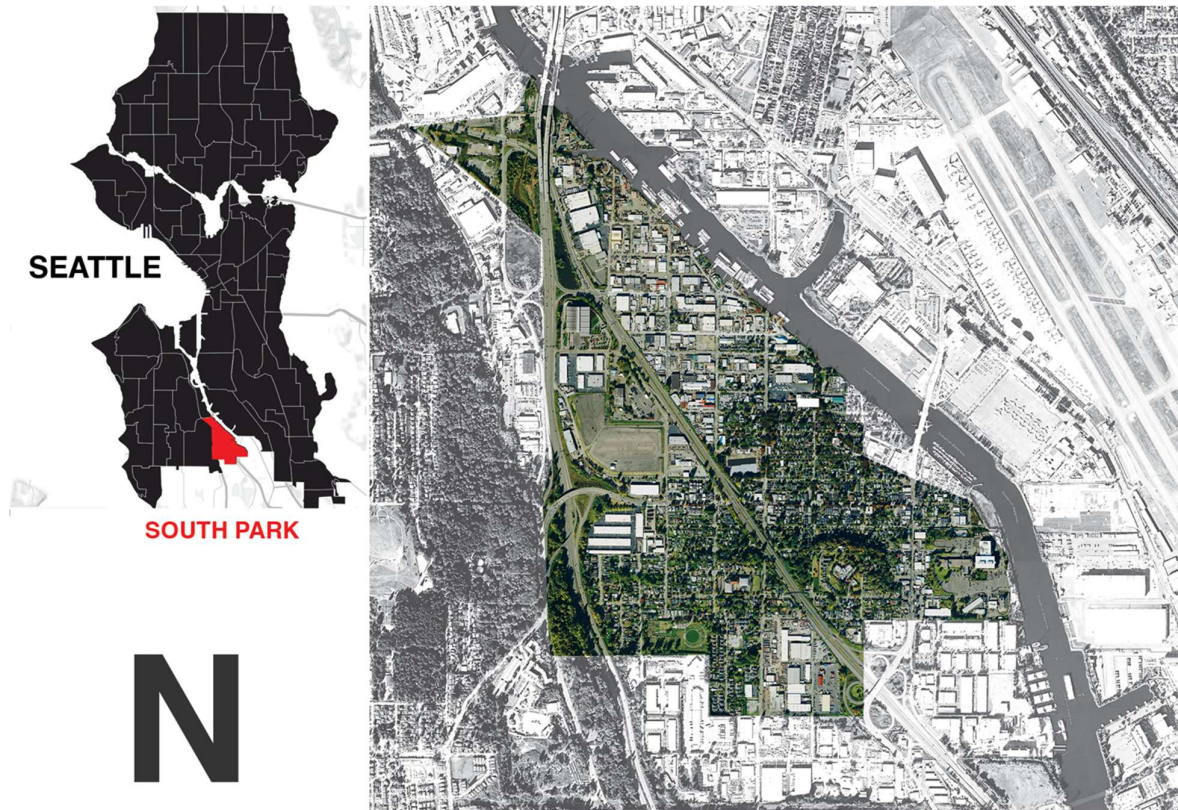


Figure 2. South Park neighborhood's boundary.

These vulnerabilities are becoming more acute as Seattle faces increasing heat risk. Because of historically mild summers, air conditioning is not widespread in Seattle. However, climate projections indicate that excessive heat events will become more frequent, longer in duration, and more severe over coming decades [61]. This trend poses disproportionate risks for low-income households, older adults, and other vulnerable populations with limited access to cooling [62,63].

The case study is also relevant from a policy perspective. Recent regulations, including the Washington State Clean Buildings Performance Standards and Seattle's Building Emissions Performance Standards (BEPS), are increasing pressure to improve energy performance and reduce fossil-fuel dependence in existing buildings [64,65]. For older multifamily and affordable housing, these requirements create both a compliance challenge and an opportunity to align decarbonization investments with indoor environmental quality and occupant health benefits.

2.1.2. Archetype Development

The first stage of the methodology focuses on constructing representative residential building archetypes that capture the dominant structural and operational characteristics of the housing stock in Seattle's Duwamish Valley. Archetype-based modeling enables scalable analysis of building energy performance by representing heterogeneous building stocks through a limited set of standardized building typologies. This approach allows simulation-based evaluation of retrofit strategies across neighborhood and urban scales while maintaining computational tractability.

A qualitative assessment of the regional affordable housing stock was conducted in collaboration with a major public-sector housing provider in Seattle to develop the archetype models to ensure that building characteristics accurately reflect local housing conditions. Baseline building models were informed by building stock attributes provide by Seattle Housing Authority (SHA).

Analysis of the organization's property inventory indicated that duplexes, quadplexes, and small apartment buildings constitute a substantial portion of the managed housing portfolio. Building documentation, construction vintages, unit counts, and consultations with facility management staff were used to refine representative geometric configurations for these dominant typologies. Programmatic load assumptions—such as occupancy, lighting, and plug loads—were adapted from the U.S. Department of Energy (DOE) Reference Building models for ASHRAE Climate Zone 4C (Seattle) and integrated into the local archetypes for physics-based simulation. In addition, historical weather data for Seattle—including temperature, humidity, and solar radiation profiles—were incorporated to define environmental boundary conditions for subsequent energy simulations.

To quantitatively explore the local building stock, a comprehensive dataset was assembled from residential buildings located within the study area. Building information was compiled from multiple sources, including building inventory databases, King County Tax Parcel records, census datasets, geographic information system (GIS) layers, and other publicly available municipal data. Integrating these datasets enabled the construction of a detailed representation of the physical, operational, and environmental characteristics of residential buildings within the community. To identify dominant residential building typologies, k-means clustering was used to group buildings into representative archetypes. Key attributes—including building height, building footprint, floor area, construction vintage—were extracted from the dataset and used as input variables for clustering analysis. This ensures that the archetype set adequately represents the broader residential building distribution within the study area, reflecting the prevalence of housing typologies identified in Tax Assessor records for residential parcels in South Park and the wider Duwamish Valley.

The final archetype set comprises four representative residential typologies: Single-family residence, Duplex, Quadplex, and Ten-unit apartment building. Figure 3 illustrates the axonometric representations of the four residential typologies adopted in the modeling framework. Table 1 summarizes the geometric characteristics of the archetype models, including conditioned floor area, envelope surface areas, and glazing ratios. These archetypes define the geometric basis for the physics-based energy models used in this study and serve as the foundation for the subsequent parametric simulation and machine learning analyses.

Table 1. Geometric characteristics of the archetype energy models developed from a qualitative analysis of the low-income residential building stock in Seattle.

Variables / Units	Single Family	Duplex	Quadplex	10-unit Apartment
Net Conditioned Area [ft ²]	1247	2374	3445	8120
Gross Roof Area [ft ²]	798	1587	1840	3431
Wall Area [ft ²]	1977	3096	4146	6553
Glazing Area [ft ²]	109	217	649	1187
Window-to-wall ratio (WWR) [%]	5.5%	7.0%	15.7%	18%
Number of floors (#)	2	2	2	3

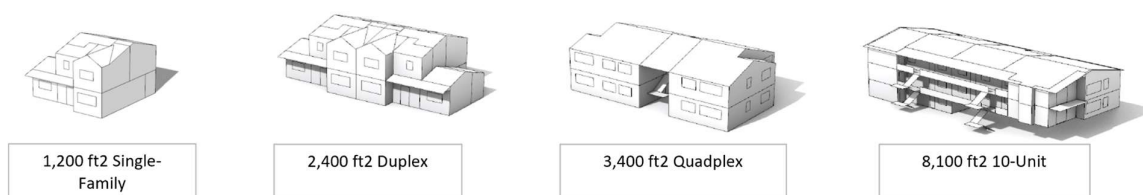


Figure 3. The four archetype models used to assess energy retrofit measures for improved human health, developed from a qualitative assessment of the low-income housing stock in Seattle.

2.2. Physics-Based Energy Modeling of Existing and Retrofit Conditions

To quantify the energy performance of representative residential building archetypes and generate training data for the machine-learning surrogate models, physics-based parametric simulation models were developed using the EnergyPlus whole-building energy simulation program. EnergyPlus was selected due to its widely validated capabilities for modeling building energy systems, thermal loads, and HVAC performance under varying operating conditions.

Table 2 summarizes the simulation parameters, which span a range of envelope, glazing, infiltration, heating system, ventilation, and domestic hot water system characteristics. These parameters reflect conditions typical of the existing Seattle housing stock, as well as retrofit scenarios that either align with current Seattle energy code requirements or represent interventions associated with improved human health outcomes (as described in Section 1.1). Specifically, the modeled variables include: 1) the extent of insulation in the wall, roof, and glazing (glazing insulation amount is represented by the total conductance U-Factor of the glazing assembly), 2) the infiltration rate (representing the air tightness of the building envelope), 3) heating system type (electric, gas furnace, and heat pump systems with various coefficients of performances (COPs)), 4) ventilation system type (either continuous exhaust typical of older construction or ERVs), and 5) hot water system type (electric resistance, gas, or heat pump).

Table 2. Energy Model Parameters Studied to Assess Relative Impact of Retrofit Interventions on Energy Consumption.

Variable Inputs	Values	Additional Description
Massing (4 options)	1) Single Family 2) Duplex 3) Quadplex 4) 10-unit	See Figure 3 for massing
Wall Insulation Value (3 options)	1) 10 ft ² ·°F·h/Btu 2) 15 ft ² ·°F·h/Btu 3) 20 ft ² ·°F·h/Btu	Low insulation performance (circa 1980-2012) Medium insulation performance Advanced insulation performance (Seattle energy code)
Roof Insulation Value (3 options)	1) 17 ft ² ·°F·h/Btu 2) 37 ft ² ·°F·h/Btu 3) 47 ft ² ·°F·h/Btu	Low insulation performance (circa 1980-2004) Medium insulation performance Advanced insulation performance (Seattle energy code)
Window Assembly U-Factor (3 options)	1) 0.57 Btu/ft ² ·°F·h 2) 0.35 Btu/ft ² ·°F·h 3) 0.29 Btu/ft ² ·°F·h	Low performing, old double-pane Typical double-pane High performance double-pane
Infiltration Rate (3 options)	1) 0.00055 ft ³ /s per ft ² of façade 2) 0.00045 ft ³ /s per ft ² of façade 3) 0.00035 ft ³ /s per ft ² of façade	Baseline from DOE reference building (low performance) Mid performance envelope High performance (Seattle code requirement)

Heating/Cooling System (3 options)	1) Electric Resistance	COP \approx 1, no cooling
	2) Gas Furnace	COP \approx 0.8, lowest energy performance w/ indoor combustion, no cooling
	3) Heat Pump	COP \approx 2.7, highest energy performance with mechanical cooling
Ventilation System (2 options)	1) Exhaust Fan	50 cfm, no heat exchange and no outdoor air filtration
	2) Energy Recovery Ventilator	50 cfm, 84% sensible heat exchange, MERV-13 filter
Hot Water System (3 options)	1) Electric Resistance	COP \approx 1
	2) Gas	COP \approx 0.8, lowest energy performance with indoor combustion
	3) Heat Pump	COP \approx 3, greatest energy performance and no indoor combustion

A parametric modeling workflow was implemented using Grasshopper, a visual programming tool, to develop a physics-based parametric analysis script. Ladybug 1.4, a plug-in for Grasshopper that builds and runs energy models through EnergyPlus, was used to set properties for the energy models and run the parametric simulations. This automated framework enabled the simulation of various possible combination of building attributes that correspond to various energy efficiency and human health measures across the identified residential typologies, producing a comprehensive dataset detailing energy performance outcome.

A TYM3 weather file for Seattle/Tacoma International airport was used, and energy simulation results captured the annual energy consumptions for the following end uses, including heating, cooling, interior lighting, electric equipment, fans, pumps, and hot water. Programmatic loads and schedules (for people, lights, equipment, etc.) were defined by the ASHRAE Standard 90.1-2019 DOE prototype mid-rise apartment building, ensuring consistency with established building energy modeling practices.

The parametric simulation matrix generated 5,832 unique simulation scenarios across the four residential archetypes. For each simulation, annual energy consumption was recorded for each end-use category (See Table 3). The resulting simulation dataset captures the relationships between building characteristics, retrofit measures, and energy consumption outcomes. These simulation outputs form the ground-truth dataset used to train machine-learning surrogate models.

Table 3. Input and target variables for data-driven energy modeling.

Category	Variables	Unit
Building Geometry	Area	<i>sqft</i>
	WWR	Percentage (%)
Envelope properties	Wall Assembly Value	R- <i>ft² · °F/Btu</i>
	Roof Assembly Value	R- <i>ft² · °F/Btu</i>
	Window Assembly Factor	U- <i>ft² · °F/Btu</i>

		Infiltration Rate	$m^3/(s \cdot m^2)$
HVAC and Ventilation		Vent Index	Dimensionless (ERV/Exhaust Fan)
		Heating-Cooling System Index	Dimensionless (Heat Pump/Gas/Electric Resistance)
		Hot Water Type	Heat Pump / Gas / Electric
Distributed Energy Performance Targets		Heating EUJ	$kBtu/sf/yr$
		Cooling EUJ	$kBtu/sf/yr$
		Lighting EUJ	$kBtu/sf/yr$
		Electric Equipment EUJ	$kBtu/sf/yr$
		Fans EUJ	$kBtu/sf/yr$
		Pumps EUJ	$kBtu/sf/yr$
		Hot Water EUJ	$kBtu/sf/yr$
	Total Energy Performance Target		Total EUJ

2.3. Machine Learning Modeling

To enable rapid prediction of building energy performance across a wide range of retrofit scenarios, this study employs a machine-learning-based surrogate modeling framework trained on the synthetic dataset generated in Section 2.2. Within this framework, machine learning models approximate the input-output relationships produced by the physics-based energy simulations, allowing energy performance to be predicted without repeatedly executing computationally intensive simulations. This hybrid approach combines the physical fidelity of simulation-based modeling with the computational efficiency of data-driven methods, making it suitable for large-scale applications.

The dataset used for model development consists of building characteristics and retrofit parameters as input variables and energy performance metrics as target variables. Input features include building geometry, envelope properties, HVAC and ventilation configurations, and domestic hot water system types. The target output includes total EUJ. Table 3 summarizes the predictor and target variables used in the machine learning model.

Before proceeding with model training, a preliminary statistical analysis was conducted to ensure data integrity, detect potential multicollinearity, and identify relationships between input variables and total EUJ. A Variance Inflation Factor (VIF) assessment was conducted to detect potential multicollinearity among independent variables. A VIF threshold of 5 was used to identify variables that exhibited excessive collinearity, which could undermine model stability and interpretability. Features exceeding this threshold were flagged for possible removal or transformation. Following these preliminary checks, data preprocessing steps were applied, including missing value imputation and numerical feature scaling to ensure uniformity in variable distributions.

The dataset was subsequently partitioned into training and testing subsets using a 70/30 ratio split to evaluate model generalizability. To mitigate overfitting and improve model robustness, a five-fold cross-validation technique was employed which has been identified as an effective method [66], allowing models to be validated across multiple iterations.

A range of machine learning algorithms was evaluated to identify the most effective predictive model for energy performance estimation. These algorithms include both statistical methods and advanced machine learning approaches. Multiple linear regression model (MLR) [67] was used as a

baseline to assess the presence of linear relationships between input variables and energy performance. Decision tree (DTREE) [68] was included for their ability to capture non-linear relationships, while random decision forest (RDF) [91], and gradient boosting machine (GBM) [70,71] were tested as ensemble-based methods known for reducing overfitting and improving accuracy. Support vector machines (SVM) using Radial Cost function, were employed to analyze high-dimensional relationships in the dataset, and K-nearest neighbors (KNN) was used as a distance-based algorithm to estimate energy demand based on similarity to other data points. Additionally, ANNs [72], specifically multi-layer perception (MLP) with a single hidden layer testing two optimization algorithms including the Broyden–Fletcher–Goldfarb–Shanno (BFGS) [73] and Port [74] optimization algorithms, were included to capture complex non-linear dependencies in the data.

Each algorithm was trained on the dataset, and hyperparameters were optimized using grid search procedure to enhance predictive performance. Each parameter is adjusted within a defined range, ensuring a thorough and unbiased evaluation, which is crucial for both selecting the best-performing algorithm and fine-tuning it for predictive modeling. Considering the dataset's size (9 features, 5,832 observations), hyperparameters are initialized based on standard heuristics, with predefined ranges set for optimization before executing the models. Table 4 outlines each algorithm's hyperparameter values and ranges, ensuring a balanced trade-off between model accuracy and computational efficiency, ultimately optimizing performance. RStudio and the 'caret' package [75,76] are used for model development and computational tasks.

Table 4. Key hyperparameters (HP), their respective ranges, and intervals for optimizing ML models, tailored to a dataset with 9 features and 5,832 observations.

Algorithm	HP-1	Range @ Interval	HP-2	Range @ Interval	HP-3	Range @ Interval
DTREE	Min. Samples Split	[10–30] @ 5	Max Depth	[5-15] @ 2	Complexity Parameter	0.01, 0.001, 0.01, 0.05, 0.1}
RDF	Max Features	10 [10, 24] (1/5 to 1/2 of features) @ 1	Node Size	10 [5, 30] @ 1	Num. of Trees	100
GBM	Learning Rate	0.1 [0.01, 0.3] @ 0.05	Interaction Depth	2 [1–10] @ 3	Num. of Trees	100
SVM	Cost (C)	0.01 {0.001, 0.01, 0.1}	Sigma (Kernel Parameter)	0.05 [0.01, 1] @ 0.05	N/A	N/A
k-NN	Num. of Neighbors (K)	3 [3, 10] @ 1	Distance Metric	Euclidean (Fixed)	Weighting	Uniform (Fixed)

ANN	Num. of Hidden Neurons	1–9 (~2/3 X Features) @ 1	Num. of Max Iterations	1000 [100 – 10000] @ 100	Weight Decay	0.5 {0.01, 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9}
-----	------------------------	---------------------------	------------------------	--------------------------	--------------	--

Model performance was evaluated using several statistical metrics commonly applied in building energy prediction studies. The coefficient of determination (R^2) was used to measure the proportion of variance in the target variable explained by the model. Mean squared error (MSE) and root mean squared error (RMSE) were calculated to quantify the average deviation between predicted and observed values, with RMSE providing an interpretable measure in the same units as energy use intensity. Additionally, mean absolute error (MAE) was employed to evaluate the absolute differences between predicted and actual values, offering a robust measure of overall model performance. These metrics were used to compare model accuracy, with an emphasis on maximizing R^2 and minimizing RMSE and MSE to identify the most reliable predictive approach. The formulations of these metrics are provided in Equations (1)–(4).

$$R^2 = 1 - \frac{\sum_{i=1}^k (y_{\text{pred},i} - y_{\text{act},i})^2}{\sum_{i=1}^k (y_{\text{act},i} - \bar{y}_{\text{act}})^2} \quad (1)$$

$$\text{MAE} = \frac{1}{k} \sum_{i=1}^k |y_{\text{pred},i} - y_{\text{act},i}| \quad (2)$$

$$\text{MSE} = \frac{1}{k} \sum_{i=1}^k \frac{\sum_{i=1}^k (y_{\text{pred},i} - y_{\text{act},i})^2}{k} \quad (3)$$

$$\text{RMSE} = \frac{1}{k} \sum_{i=1}^k \sqrt{\frac{\sum_{i=1}^k (y_{\text{pred},i} - y_{\text{act},i})^2}{k}} \quad (3)$$

where, $y_{\text{act},i}$ and $y_{\text{pred},i}$ as actual and predicted values, \bar{y}_{act} as mean of the actual values, and n as total number of observations.

Following model comparison, the algorithm demonstrating the highest predictive accuracy and generalization capability was selected for further analysis and interpretation. To better understand model behavior and identify the most influential predictors of building energy consumption, two model interpretation techniques were applied. Variable importance analysis was performed using the Garson algorithm [77], which has been widely used to quantify the relative contribution of input variables in neural network models [78]. In addition, Partial Dependence Plots (PDPs) [71] were used to explore the interactions between key predictors and energy load estimates. PDPs are valuable for interpreting ML models since they illustrate the average effect of one or more features on the predicted outcome, while holding all other features constant. This allows us to visualize how changes in a particular predictor influence model prediction, even when dealing with complex, non-linear, or “black-box” models. Additionally, PDPs help uncover trends, interactions, and potential thresholds in the data, offering clearer insights into the model’s behavior and decision-making process.

3. Results

3.1. Archetype Development and Physics-Based Simulation

Four representative residential archetypes were derived to reflect the dominant housing typologies in Seattle’s Duwamish Valley: a single-family house, duplex, quadplex, and ten-unit apartment building. These archetypes span the principal geometric and operational characteristics of the local residential stock and form the basis of the physics-based simulation framework (Table 2; Figure 3). Using these archetypes, the parametric simulation workflow generated 5,832 unique

simulation cases across variations in envelope performance, infiltration, HVAC systems, ventilation strategy, and domestic hot water systems.

3.2. Machine Learning Model Performance and Selection

The predictive performance of the evaluated machine learning models is summarized in Table 5. Model accuracy was assessed using R^2 , MSE, RMSE, and MAE for both training and testing datasets under a 70/30 train–test split. Among the evaluated models, RDF achieved the highest overall predictive accuracy, with $R^2 = 0.98$ for both training and testing datasets and the lowest prediction errors (RMSE = 2.49 for training and 2.57 for testing). GBM and SVM also performed strongly, each achieving $R^2 = 0.95$ on the test set. The ANN model demonstrated similarly strong performance, with $R^2 = 0.94$, RMSE = 3.97, and MAE = 2.67 on the testing dataset. In contrast, DTREE and k-NN showed lower predictive performance, with k-NN exhibiting the largest drop in test accuracy and the highest generalization error.

Table 5. Comparative performance metrics of ML models for predicting energy use: training versus test set evaluation.

Train Set					Test Set			
Algorithm	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE
MLR	0.62	-	-	-	0.63	-	-	-
DTREE	0.81	48.86	6.99	5.51	0.82	47.14	6.87	5.43
RDF	0.98	6.22	2.49	1.75	0.98	6.60	2.57	1.81
GBM	0.95	21.52	4.64	3.38	0.95	21.11	4.59	3.35
SVM	0.95	16.94	4.12	2.64	0.95	17.32	4.16	2.71
k-NN	0.89	27.76	5.27	4.01	0.84	43.16	6.57	5.05
ANN	0.94	15.42	3.93	2.72	0.94	15.73	3.97	2.67

Comparison of training and testing performance indicates that RDF, GBM, SVM, and ANN all generalized well, with only minor reductions in performance between datasets. RDF showed the smallest difference between training and testing RMSE, indicating minimal overfitting. ANN also maintained stable predictive performance across both datasets, supporting its robustness as a surrogate model.

Despite RDF achieving the highest R^2 , ANN was selected as the primary model for further analysis due to its strong predictive accuracy and ability to capture complex nonlinear relationships among retrofit variables. The final ANN MLP configuration employed the Port optimization algorithm, which provided stable convergence and improved model generalization across the simulated parameter space. Figure 4 illustrates the architecture of the ANN model. Predicted values closely matched simulated outputs for both the training and testing datasets (Figure 5), indicating robust model generalization. Unlike tree-based models that partition the feature space, ANN's weight-based learning captures continuous nonlinear dependencies that are critical for understanding variations in building energy use. Additionally, the lower MAE achieved by the ANN model suggests more stable predictions, which is particularly important for retrofit-driven energy efficiency analysis requiring reliable estimates at the individual building level.

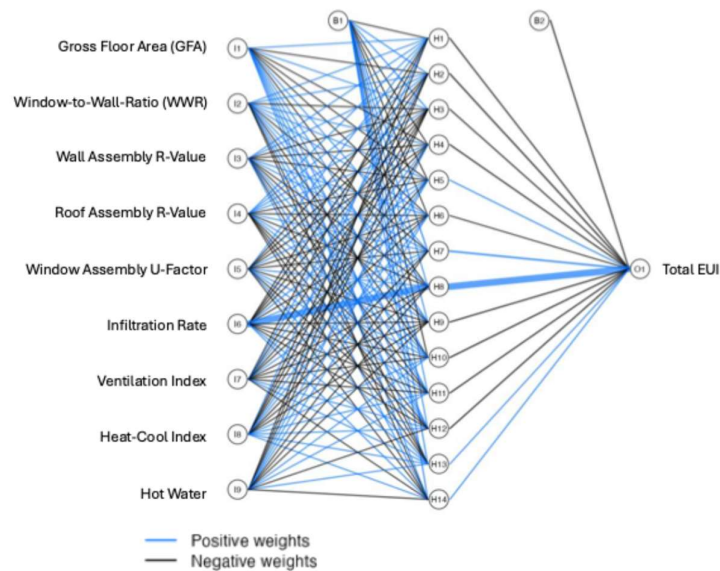


Figure 4. Architecture of the ANN used for Energy Use Prediction. The figure illustrates the structure of the ANN model, depicting the connections between input features (building parameters) and the total energy use intensity (EUI). The network consists of multiple hidden layers, where blue lines represent positive weights, and black lines indicate negative weights. The thickness of the lines corresponds to the strength of the connection, highlighting the influence of each input variable on the final energy consumption prediction.

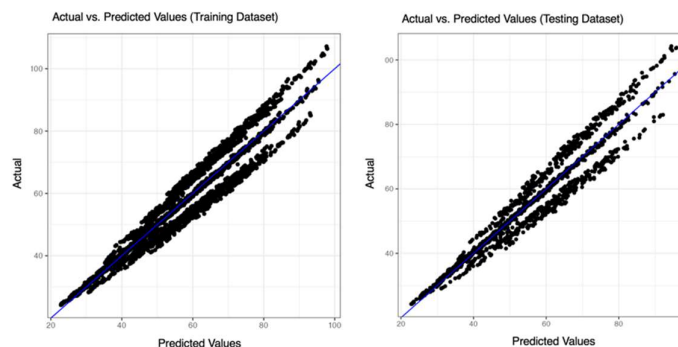


Figure 5. Performance of the ANN surrogate model. Actual vs. predicted values for the training (left) and testing (right) datasets. The blue line represents the 1:1 reference line, showing strong agreement between predicted and simulated values.

3.3. Key Drivers of Energy Performance and Health-Driven Retrofits

Variable importance analysis and PDPs were used to identify the variables that most strongly influence building energy performance and to interpret their effects and their implications for health-driven energy retrofits within the modeling framework. Figure 6 shows the relative importance of input variables estimated using the Garson algorithm, while Figure 7 presents selected PDPs illustrating key predictor–response relationships.

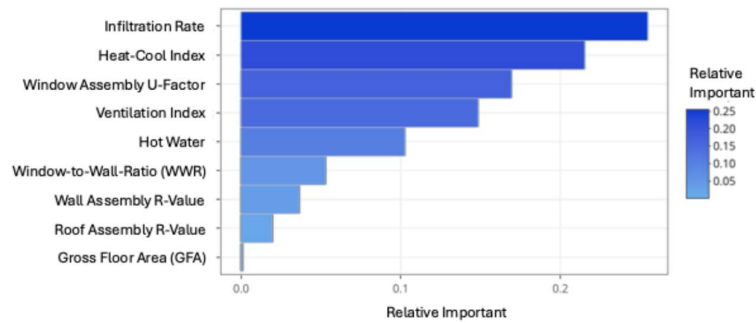


Figure 6. Relative variable importance analysis using the Garson algorithm, highlighting the relative influence of different building characteristics on energy performance.

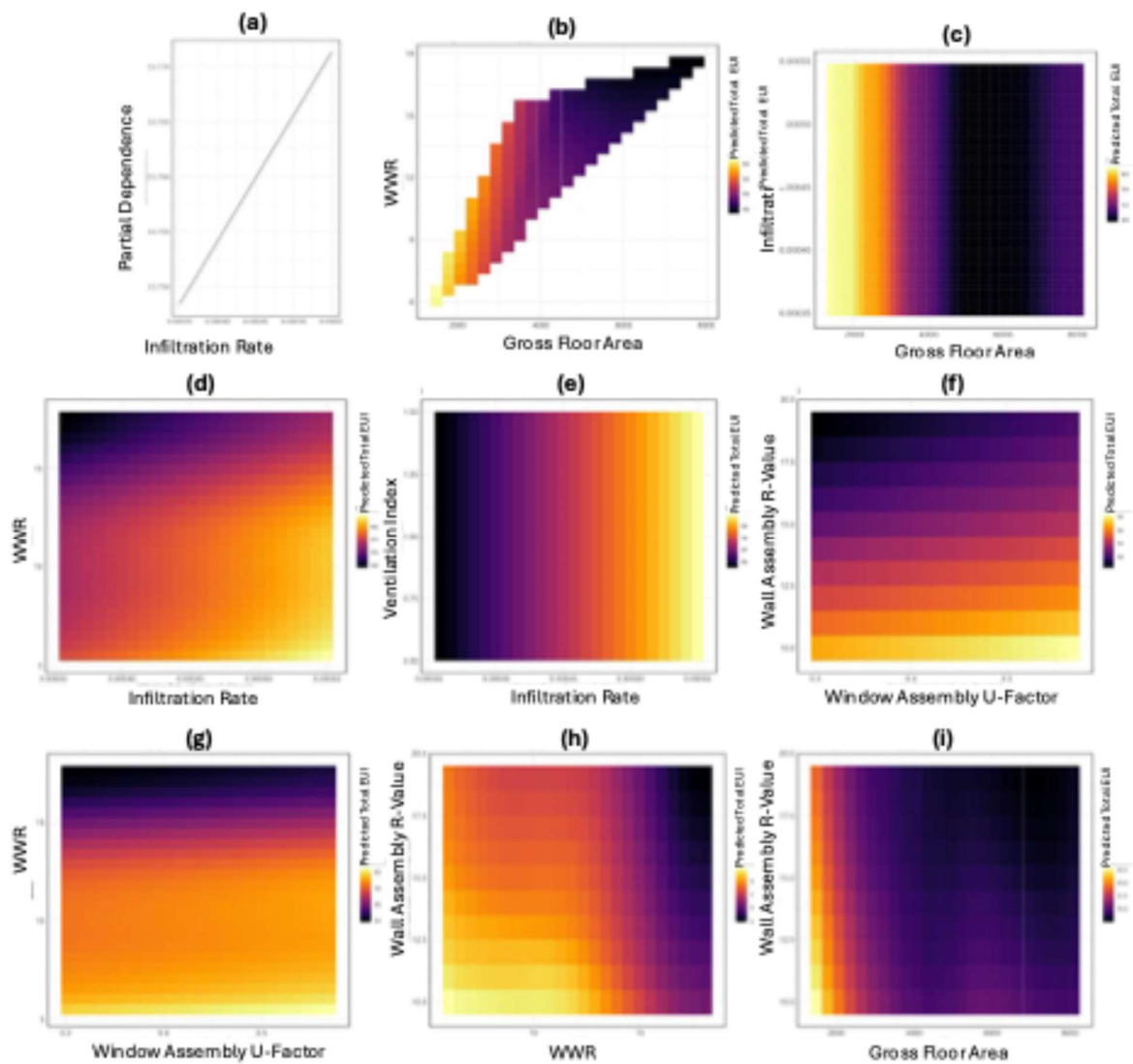


Figure 7. PDPs illustrating interactions between key building characteristics in predicting energy performance.

Variable importance analysis (Figure 6) identifies infiltration rate as the dominant predictor of total energy use, followed by the heating–cooling system index, indicating that envelope airtightness and HVAC efficiency are the strongest determinants of modeled energy performance. Window U-

factor and ventilation system type also contribute meaningfully, whereas wall and roof insulation levels have smaller relative influence within the evaluated parameter ranges. WWR has moderate importance, while gross floor area appears less influential once envelope and system characteristics are considered. Together, these results suggest that the highest-value retrofit measures are those that improve airtightness, mechanical efficiency, and selected aspects of envelope performance.

To further interpret the model behavior, PDPs (Figure 7) were used to examine the marginal influence of key predictors on predicted energy use. Because PDPs represent model-averaged responses across the sampled feature space, the gradients shown in Figure 7 reflect average model responses within the observed data distribution rather than direct causal effects of changing individual parameters in isolation.

The one-dimensional PDP for infiltration (Figure 7a) shows a monotonic and approximately linear increase in predicted EUI as infiltration raises across the sampled range ($\approx 3.5 \times 10^{-4}$ – 5.5×10^{-4}). Predicted EUI increases slightly from approximately 53.748 to 53.772 across this interval, indicating that, after marginalizing over the distribution of all other inputs, the trained model consistently penalizes higher air leakage, with no evident threshold or curvature. Because this curve represents a model-averaged intervention response rather than an empirical correlation, it can be interpreted as the learned marginal sensitivity of performance to envelope airtightness under controlled variation of infiltration while integrating over other covariates [71,79]. In retrofit terms, the small magnitude over this narrow leakage band suggests air-sealing benefits may be most pronounced for buildings with leakage rates outside the modeled range and/or when combined with complementary envelope measures.

The two-dimensional PDP for GFA and WWR (Figure 7b) indicates that predicted EUI ranges from approximately 43.5 to 61.5 within the data-supported domain (GFA $\approx 1,200$ – $8,000$ ft²; WWR ≈ 6 – 18%). The highest predicted EUI occurs in small buildings with low WWR, while lower values occur in larger buildings with higher WWR, corresponding to a reduction of roughly 15–18 EUI units across the surface. These results suggest that smaller buildings exhibit higher energy intensity and building size can serve as an important segmentation variable when prioritizing retrofit interventions.

The PDP examining GFA and infiltration rate (Figure 7c) further demonstrates that, the predicted EUI spans approximately 48 to ~ 61.5 and is dominated by variation along GFA. Predicted EUI is highest at GFA $\approx 1,200$ – $2,000$ (≈ 60 – 61.5) across essentially the full infiltration range, declines through mid-size buildings to the minimum band (~ 48 – 50) around GFA $\approx 5,000$ – $6,500$, and then increases slightly at the largest sizes (GFA $\approx 7,500$ – $8,500$) to roughly ~ 50 – 52 , yielding a non-monotonic size pattern with an overall reduction. In contrast, variations in infiltration within the tested range produce comparatively small changes in predicted EUI at fixed GFA. These results suggest that building size exerts a stronger influence on energy intensity than infiltration within the sampled leakage band, reinforcing the importance of size-based segmentation when evaluating retrofit strategies.

Interactions between infiltration and envelope characteristics further highlight the importance of airtightness improvements. The PDP for infiltration and WWR (Figure 7d) indicates that, the predicted EUI increases strongly with infiltration across the WWR range. At lower WWR values (≈ 5 – 10%), increasing infiltration from the lower to upper bound of the sampled range raises predicted EUI approximately $\approx +4$ – 6 units, with the global maximum occurring at the high-infiltration/low-WWR region. Increasing WWR tends to reduce predicted EUI within the modeled domain, but infiltration remains the dominant driver of variation. These relationships should therefore be interpreted as model-based sensitivities within the sampled design space rather than direct causal effects of changing WWR or infiltration in isolation. These results emphasize the importance of air-sealing interventions, particularly in buildings with higher leakage rates.

Similarly, the PDP for infiltration and ventilation index (Figure 7e) indicates a predominantly additive, monotonic response over infiltration and ventilation index, with the model's predicted EUI spanning approximately 55.5 to 59.8 and varying almost entirely along the infiltration axis. Changes in ventilation index produce only minor variations in predicted EUI when infiltration is held

constant. Within the modeled domain and sampled ranges, the PDP implies minimal interaction between the two variables and suggests that reducing uncontrolled infiltration provides substantially larger energy benefits than modifying ventilation intensity alone. The model prioritizes airtightness improvements over ventilation-intensity changes as the higher-leverage intervention within this feature pair and domain.

The PDP examining window assembly U-factor ≈ 0.30 – 0.58 and wall assembly R-value ≈ 10 – 20 (Figure 7f) indicates that, the model's predicted EUI spans approximately ~ 51.5 to ~ 56 and exhibits a predominantly additive, monotonic structure with limited interaction between the two variables. Increasing wall insulation from approximately $R \approx 10$ to $R \approx 20$ reduces predicted EUI by roughly ~ 3.5 – 4.5 EUI units, while increasing U-factor within the sampled range produces a smaller increase in predicted EUI (~ 0.5 – 1.5 units). From a retrofit-planning perspective, the results suggest prioritizing wall-insulation upgrades where R is low, and treating window U-factor improvements as a secondary measure with diminishing marginal impact in highly insulated-wall cases which provide greater energy benefits.

The PDP for window U-factor and WWR (Figure 7g) shows that predicted EUI varies more strongly along WWR axis than along the U-factor axis within the sampled domain. Increasing WWR from ~ 5 – 6% to ~ 16 – 17% corresponds to a reduction of roughly 5–6 EUI units whereas changes in U-factor produce relatively small variations at fixed WWR. For retrofit planning, because WWR is generally not a practical retrofit lever and the PDP summarizes model-averaged responses rather than prescribing window-area changes, the actionable use of this surface is to treat WWR as a segmentation variable when prioritizing glazing upgrades: within $WWR \approx 5$ – 10 , improving window thermal performance across the evaluated U range corresponds to a modest modeled reduction ($\leq \sim 0.5$ – 1 EUI unit), whereas at $WWR \approx 16$ – 17 the model indicates little additional benefit from U-factor improvements alone, suggesting that window-performance retrofits are most likely to be differentiated (in modeled impact) in the lower-WWR segment of the stock.

The PDP for WWR and wall assembly R-value (Figure 7h) indicates that increasing wall R-value from approximately 10 to 20 reduces predicted EUI by about 3–5 units in low-WWR buildings. The result implies that increasing wall R-value delivers the largest modeled reductions in the low-WWR regime (~ 3 – 5 units) but yields limited incremental benefit at high WWR (~ 0 – 2 units), so wall-insulation retrofits are most likely to be differentiated (in modeled impact) in buildings with lower-to-moderate WWR within the evaluated domain.

The PDP for GFA and wall assembly R-value (Figure 7i) shows that the predicted EUI ranges from approximately ~ 54 to ~ 63 . The result indicates that predicted EUI decreases strongly with increasing building size, ranging from approximately 62–63 in small buildings to about 54–55 in larger buildings. For retrofit, this results show that wall-insulation upgrades are most likely to yield the largest modeled reductions in the low-GFA segment (~ 2 – 4 units at $GFA \approx 1,200$ – $2,000$), whereas large-floor-area buildings ($GFA \geq 6,000$) show limited incremental modeled benefit ($\leq \sim 1$ unit) from additional wall R within the evaluated range, suggesting insulation retrofits may be most impactful when targeted to smaller buildings with lower existing envelope performance

Synthesizing the PDP results (Figures 7a–7i), the largest modeled energy-reduction potential occurs in smaller buildings with weaker envelope performance, where energy intensity is highest within the evaluated domain. Across multiple feature combinations, infiltration reduction consistently emerges as the most influential actionable measure, followed by HVAC system upgrades and targeted envelope improvements.

4. Discussion

The results demonstrate that the most influential retrofit variables in the modeled residential building stock are those that affect both energy demand and indoor environmental conditions, particularly infiltration rate, HVAC system type, window thermal performance, and ventilation strategy. This is significant because it suggests that energy-efficiency retrofits in disadvantaged urban communities can be framed not only as climate mitigation strategies but also as interventions with

health co-benefits. In the present study, the variables associated with the largest modeled reductions in energy use are also closely linked to thermal stability, indoor air quality, and exposure to outdoor pollutants, supporting the concept of health-driven energy efficiency within urban retrofit planning.

Among all predictors, infiltration rate emerged as the dominant driver of modeled energy performance, and the PDP analysis shows that higher infiltration consistently increases predicted EUI. This trend is consistent with building physics in heating-dominated climates such as Seattle, where infiltration heat loss scales with air exchange rate and indoor–outdoor temperature differences ($Q = \rho c_p V \Delta T$), and with field and simulation evidence that heating penalties can remain approximately proportional to leakage over low-to-moderate airtightness ranges [80,81]. Previous studies estimate that air leakage can account for 25–40% of residential heating and cooling energy use [82], while tighter building envelopes are associated with improved IAQ, indoor comfort, and reduced drafts [83]. In environmentally burdened communities such as Seattle’s Duwamish Valley—where residents face elevated exposure to traffic-related and industrial pollution—reducing uncontrolled infiltration can therefore provide dual benefits by lowering energy demand while limiting the ingress of outdoor pollutants and allergens. However, airtightness improvements should be implemented alongside controlled mechanical ventilation to maintain adequate indoor air quality.

The heat-cool index was identified as another dominant predictor of energy performance highlighting the importance of HVAC system efficiency in shaping building energy demand. Replacing conventional heating systems—such as electric resistance heating or gas furnaces—with high-efficiency heat pumps can substantially reduce operational energy use while improving thermal comfort and health resilience. Previous studies show that heat pump systems can significantly reduce building energy consumption and greenhouse gas emissions while supporting building electrification strategies [84–86]. In addition to energy savings, heat pumps also improve indoor thermal stability and eliminate indoor combustion sources, which can reduce exposure to pollutants associated with gas appliances. From a health perspective, efficient cooling capacity is increasingly important as extreme heat events become more frequent due to climate change, with studies linking inadequate indoor cooling to higher risks of heat-related illness and mortality [86–89]. These findings highlight the importance of HVAC electrification as both an energy-efficiency and public-health intervention.

The role of ventilation systems should be interpreted alongside infiltration results. Although the ventilation index contributes to model predictions, the PDP analysis indicates that within the sampled parameter range ventilation variations produce smaller changes in predicted EUI compared with infiltration. This suggests that ventilation is not a primary driver of energy intensity but remains essential for indoor environmental quality. Mechanical ventilation systems such as energy recovery ventilators (ERVs) can provide fresh air while limiting energy losses, particularly in airtight buildings. Adequate ventilation is also essential for reducing indoor pollutant concentrations—including particulate matter (PM_{2.5}), carbon dioxide (CO₂), and volatile organic compounds (VOCs)—that are associated with respiratory illness, cognitive impairment, and reduced productivity [90,91] [92]. The adaptive ventilation strategies such as demand-controlled ventilation (DCV) are identified as a cost effective and energy efficient strategy [93] that reduces ventilation-related energy use while maintaining good IAQ [94] particularly in areas affected by seasonal air pollution or wildfire events. Implementing ERVs ensures a balance between fresh air supply and energy conservation, particularly in airtight buildings, where mechanical ventilation is necessary to maintain IAQ while avoiding excessive energy use. This approach is crucial during airborne disease outbreaks or wildfire seasons, when opening windows is not a viable option. Consequently, the results support a combined strategy of infiltration reduction with controlled mechanical ventilation, rather than treating these measures as substitutes.

Envelope characteristics also influence predicted energy performance. Window thermal performance (U-factor) was identified as an important predictor, while the PDP analysis suggests that wall insulation improvements often produce larger reductions in predicted EUI than window upgrades within the evaluated parameter ranges, particularly in smaller buildings or buildings with

lower existing insulation levels. Poor-performing fenestration systems increase conductive heat transfer and are widely recognized as contributors to higher heating and cooling loads and reduced thermal comfort [95–97]. While high-performance windows—such as low-emissivity, multi-pane assemblies—can reduce heat transfer and improve indoor comfort, the modeling results indicate that glazing upgrades may provide smaller marginal benefits compared with wall insulation improvements in certain building segments. This suggests that envelope retrofit strategies should be tailored to building conditions rather than applied uniformly across the housing stock.

The results further show that the GFA and WWR primarily function as segmentation variables rather than direct retrofit levers. Within the data-supported domain, smaller buildings exhibit higher predicted EUI and show stronger modeled benefits from infiltration reduction and insulation improvements. This pattern aligns with national housing datasets such as the U.S. Residential Energy Consumption Survey, which indicate that smaller housing units often exhibit higher energy use per unit floor area because certain end uses—such as appliances and plug loads—do not scale proportionally with floor area. In contrast, WWR is typically fixed in existing buildings and should therefore be interpreted as a contextual design variable that influences the effectiveness of other envelope retrofits rather than as a direct retrofit parameter.

Taken together, the results suggest a retrofit prioritization hierarchy in which air-sealing measures are implemented first, followed by HVAC system upgrades, particularly high-efficiency heat pumps, and then targeted envelope improvements, especially wall insulation in smaller and more weakly insulated buildings. Window performance improvements remain beneficial but appear to provide smaller marginal gains within the tested parameter ranges. Importantly, this hierarchy aligns energy-efficiency priorities with health-relevant indoor environmental outcomes, including reduced pollutant ingress, improved thermal stability, and better ventilation control. In this way, the modeling framework supports a health-driven approach to residential retrofits, integrating decarbonization objectives with indoor environmental quality and community health considerations.

This study also contributes methodologically by demonstrating the potential of machine-learning surrogate modeling integrated with physics-based simulations for urban retrofit analysis. The ANN-based surrogate model captures nonlinear interactions among building variables while enabling rapid evaluation of large retrofit scenario spaces. This approach is particularly valuable for urban building stocks, where exhaustive simulation of all retrofit combinations can be computationally prohibitive. The use of feature importance analysis and partial dependence plots further improves interpretability, allowing the surrogate model to provide insights that can inform retrofit prioritization and policy decisions.

Several limitations should be acknowledged. First, the modeling framework relies on synthetic data generated from parametric simulations, rather than long-term measured operational data. Consequently, the results represent typical operating conditions defined by the simulation assumptions and training dataset. Second, the framework focuses on energy performance and health-relevant indoor environmental proxies, rather than directly modeling indoor pollutant concentrations or occupant health outcomes. Therefore, the health implications discussed here should be interpreted as inferred co-benefits supported by existing literature, rather than direct measurements of health impacts. Third, the PDP results represent model-averaged responses within the sampled feature space, and should not be interpreted as causal relationships outside the data-supported domain.

Future research should integrate measured operational data, indoor environmental sensing, and occupant-centered indicators to strengthen model calibration and improve real-world applicability. While direct modeling of occupant health or comfort outcomes is beyond the scope of this work, future integration of occupant-centric data [33,98]—such as comfort metrics, behavioral patterns obtained from the environmental and social sensing [99–103], microclimate data [104] or real-time IAQ monitoring [105,106]—would provide a stronger empirical basis for linking retrofit scenarios to human-building-environment relations and health outcomes. Such integration would support more

precise targeting of interventions and help develop retrofit strategies that simultaneously advance energy efficiency, environmental justice, and public health in vulnerable urban communities.

5. Conclusion

This study presents an AI-enhanced physics-based modeling framework for evaluating residential retrofit strategies that simultaneously improve energy performance and indoor environmental conditions linked to occupant health. By integrating parametric energy simulations with machine-learning surrogate modeling, the framework enables efficient exploration of retrofit scenarios across representative residential archetypes in Seattle's Duwamish Valley. The results identify infiltration rate, HVAC system efficiency, window thermal performance, and ventilation strategy as key drivers of building energy intensity, with infiltration reduction and HVAC electrification emerging as the most impactful interventions within the modeled domain. Envelope improvements, particularly wall insulation in smaller and more weakly insulated buildings, provide additional benefits, while glazing upgrades offer more targeted improvements depending on building characteristics. Importantly, these retrofit measures generate co-benefits for indoor environmental quality by reducing pollutant ingress, improving thermal stability, and enhancing ventilation control, thereby supporting a health-driven approach to residential energy efficiency. Methodologically, the study demonstrates how combining physics-based simulations with interpretable machine-learning models can support scalable retrofit prioritization and decision-making for urban building stocks. The proposed framework contributes a transferable approach for advancing equitable and health-oriented building decarbonization strategies in environmentally burdened communities.

Acknowledgments: This research was supported by grants from the University of Washington Population Health Initiative (PHI), including the Climate Change Pilot Grant and the Tier 3 Pilot Research Grant for the project "*DecarbCityTwin: A Platform for Health-Driven and Equitable Decarbonization of the Built Environment.*" The authors gratefully acknowledge the partnership and collaboration of the City of Seattle Office of Sustainability and Environment and the Duwamish River Community Coalition (DRCC). The content of this publication is solely the responsibility of the authors and does not necessarily reflect the views of the supporting organizations.

References

1. Ashayeri M, Abbasabadi N. A framework for integrated energy and exposure to ambient pollution (iEnEx) assessment toward low-carbon, healthy, and equitable cities. *Sustainable Cities and Society* 2022;78:103647. <https://doi.org/10.1016/j.scs.2021.103647>.
2. Bullard RD, Mohai P, Saha R, Wright B. TOXIC WASTES AND RACE AT TWENTY: WHY RACE STILL MATTERS AFTER ALL OF THESE YEARS. *Environmental Law* 2008;38:371–411.
3. Bednar DJ, Reames TG. Recognition of and response to energy poverty in the United States. *Nature Energy* 2020;5:432–9. <https://doi.org/10.1038/s41560-020-0582-0>.
4. Mendez M, Blond N, Amedro D, Hauglustaine DA, Blondeau P, Afif C, et al. Assessment of indoor HONO formation mechanisms based on in situ measurements and modeling. *Indoor Air* 2017;27:443–51. <https://doi.org/10.1111/ina.12320>.
5. Dentz J, Conlin F, Podorson D, Alaigh K. Public Housing: A Tailored Approach to Energy Retrofits n.d.
6. Zahed F, Pardakhti A, Motlagh MS, Mohammad Kari B, Tavakoli A. Infiltration of outdoor PM_{2.5} and influencing factors. *Air Qual Atmos Health* 2022;15:2215–30. <https://doi.org/10.1007/s11869-022-01246-4>.
7. Meier R, Schindler C, Eeftens M, Aguilera I, Ducret-Stich RE, Ineichen A, et al. Modeling indoor air pollution of outdoor origin in homes of SAPALDIA subjects in Switzerland. *Environment International* 2015;82:85–91. <https://doi.org/10.1016/j.envint.2015.05.013>.
8. McCormack MC, Breyse PN, Matsui EC, Hansel NN, Peng RD, Curtin-Brosnan J, et al. Indoor particulate matter increases asthma morbidity in children with non-atopic and atopic asthma. *Annals of Allergy, Asthma & Immunology* 2011;106:308–15. <https://doi.org/10.1016/j.anai.2011.01.015>.

9. Isiugo K, Jandarov R, Cox J, Ryan P, Newman N, Grinshpun SA, et al. Indoor particulate matter and lung function in children. *Science of The Total Environment* 2019;663:408–17. <https://doi.org/10.1016/j.scitotenv.2019.01.309>.
10. Woodruff TJ, Parker JD, Schoendorf KC. Fine Particulate Matter (PM_{2.5}) Air Pollution and Selected Causes of Postneonatal Infant Mortality in California. *Environ Health Perspect* 2006;114:786–90. <https://doi.org/10.1289/ehp.8484>.
11. Takaro TK, Krieger J, Song L, Sharify D, Beaudet N. The Breathe-Easy Home: The Impact of Asthma-Friendly Home Construction on Clinical Outcomes and Trigger Exposure. *Am J Public Health* 2011;101:55–62. <https://doi.org/10.2105/AJPH.2010.300008>.
12. Carlton EJ, Barton K, Shrestha PM, Humphrey J, Newman LS, Adgate JL, et al. Relationships between home ventilation rates and respiratory health in the Colorado Home Energy Efficiency and Respiratory Health (CHEER) study. *Environmental Research* 2019;169:297–307. <https://doi.org/10.1016/j.envres.2018.11.019>.
13. Manuel J. Avoiding Health Pitfalls of Home Energy-Efficiency Retrofits. *Environ Health Perspect* 2011;119. <https://doi.org/10.1289/ehp.119-a76>.
14. Holden KA, Lee AR, Hawcutt DB, Sinha IP. The impact of poor housing and indoor air quality on respiratory health in children. *Breathe* 2023;19:230058. <https://doi.org/10.1183/20734735.0058-2023>.
15. Fisk WJ. The ventilation problem in schools: literature review. *Indoor Air* 2017;27:1039–51. <https://doi.org/10.1111/ina.12403>.
16. Howden-Chapman P, Matheson A, Crane J, Viggers H, Cunningham M, Blakely T, et al. Effect of insulating existing houses on health inequality: cluster randomised study in the community. *BMJ* 2007;334:460. <https://doi.org/10.1136/bmj.39070.573032.80>.
17. Ahrentzen S, Erickson J, Fonseca E. Thermal and health outcomes of energy efficiency retrofits of homes of older adults. *Indoor Air* 2016;26:582–93. <https://doi.org/10.1111/ina.12239>.
18. Gerardi DA. Building-Related Illness. *Clinical Pulmonary Medicine* 2010;17:276–81. <https://doi.org/10.1097/CPM.0b013e3181fa1448>.
19. Abbasabadi N, Ashayeri M. Urban energy use modeling methods and tools: A review and an outlook. *Building and Environment* 2019;161:106270. <https://doi.org/10.1016/j.buildenv.2019.106270>.
20. Chen Y, Hong T, Piette MA. Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy* 2017;205:323–35. <https://doi.org/10.1016/j.apenergy.2017.07.128>.
21. Buckley N, Mills G, Reinhart C, Berzolla ZM. Using urban building energy modelling (UBEM) to support the new European Union's Green Deal: Case study of Dublin Ireland. *Energy and Buildings* 2021;247:111115. <https://doi.org/10.1016/j.enbuild.2021.111115>.
22. Keirstead J, Jennings M, Sivakumar A. A review of urban energy system models: Approaches, challenges and opportunities. *Renewable and Sustainable Energy Reviews* 2012;16:3847–66. <https://doi.org/10.1016/j.rser.2012.02.047>.
23. Reinhart CF, Cerezo Davila C. Urban building energy modeling – A review of a nascent field. *Building and Environment* 2016;97:196–202. <https://doi.org/10.1016/J.BUILDENV.2015.12.001>.
24. Sola A, Corchero C, Salom J, Sanmarti M. Simulation tools to build urban-scale energy models: A review. *Energies* 2018;11:3269–3269. <https://doi.org/10.3390/en11123269>.
25. Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews* 2009;13:1819–35. <https://doi.org/10.1016/j.rser.2008.09.033>.
26. Li Q, Jige Quan S, Augenbroe G, Pei-Ju Yang P, Brown J. Building Energy Modelling at Urban Scale: Integration of Reduced Order Energy Model With Geographical Information. 14th Conference of International Building Performance Simulation Association, Hyderabad, India: 2015, p. 190–9.
27. Nutkiewicz A, Yang Z, Jain RK. Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied Energy* 2018;225:1176–89. <https://doi.org/10.1016/j.apenergy.2018.05.023>.

28. Wiedenhofer D, Lenzen M, Steinberger JK. Energy requirements of consumption: Urban form, climatic and socio-economic factors, rebounds and their policy implications. *Energy Policy* 2013;63:696–707. <https://doi.org/10.1016/j.enpol.2013.07.035>.
29. Yun GY, Steemers K. Behavioural, physical and socio-economic factors in household cooling energy consumption. *Applied Energy* 2011;88:2191–200. <https://doi.org/10.1016/j.apenergy.2011.01.010>.
30. Dagoumas A. Modelling socio-economic and energy aspects of urban systems. *Sustainable Cities and Society* 2014;13:192–206. <https://doi.org/10.1016/j.scs.2013.11.003>.
31. Abbasabadi N, Ashayeri M. Socioeconomic determinants of public health and residential building energy use in Chicago. *Association of Collegiate Schools of Architecture* 2021;3:707–13.
32. Worthy A, Ashayeri M, Marshall J, Abbasabadi N. Bridging the simulation-to-reality gap: A comprehensive review of microclimate integration in urban building energy modeling (UBEM). *Energy and Buildings* 2025;331:115392. <https://doi.org/10.1016/j.enbuild.2025.115392>.
33. Happle G, Fonseca JA, Schlueter A. A review on occupant behavior in urban building energy models. *Energy and Buildings* 2018;174:276–92. <https://doi.org/10.1016/j.enbuild.2018.06.030>.
34. Kontokosta CE, Tull C. A data-driven predictive model of city-scale energy use in buildings. *Applied Energy* 2017;197:303–17. <https://doi.org/10.1016/j.apenergy.2017.04.005>.
35. Schiefelbein J, Rudnick J, Scholl A, Remmen P, Fuchs M, Müller D. Automated urban energy system modeling and thermal building simulation based on OpenStreetMap data sets. *Building and Environment* 2019;149:630–9. <https://doi.org/10.1016/j.buildenv.2018.12.025>.
36. Abbasabadi N, Ashayeri M, editors. *Artificial intelligence in performance-driven design: theories, methods, and tools*. First edition. Hoboken, New Jersey: Wiley; 2024.
37. Ali U, Bano S, Shamsi MH, Sood D, Hoare C, Zuo W, et al. Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy and Buildings* 2024;303:113768. <https://doi.org/10.1016/j.enbuild.2023.113768>.
38. Ashayeri M, Abbasabadi N. A Hybrid Physics-Based Machine Learning Approach for Integrated Energy and Exposure Modeling. In: Abbasabadi N, Ashayeri M, editors. *Artificial Intelligence in Performance-Driven Design*. 1st ed., Wiley; 2024, p. 57–79. <https://doi.org/10.1002/9781394172092.ch3>.
39. Abbasabadi N, Ashayeri M. Machine Learning in Urban Building Energy Modeling. In: Abbasabadi N, Ashayeri M, editors. *Artificial Intelligence in Performance-Driven Design*. 1st ed., Wiley; 2024, p. 31–55. <https://doi.org/10.1002/9781394172092.ch2>.
40. Abbasabadi N, Ashayeri M, Azari R, Stephens B, Heidarinejad M. An integrated data-driven framework for urban energy use modeling (UEUM). *Applied Energy* 2019;253:113550. <https://doi.org/10.1016/j.apenergy.2019.113550>.
41. Cheng X, Khomtchouk B, Matloff N, Mohanty P. *Polynomial Regression As an Alternative to Neural Nets* 2018.
42. Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews* 2009;13:1819–35. <https://doi.org/10.1016/j.rser.2008.09.033>.
43. Park SK, Moon HJ, Min KC, Hwang C, Kim S. Application of a multiple linear regression and an artificial neural network model for the heating performance analysis and hourly prediction of a large-scale ground source heat pump system. *Energy and Buildings* 2018;165:206–15. <https://doi.org/10.1016/j.enbuild.2018.01.029>.
44. Papadopoulos S, Azar E, Woon W-L, Kontokosta CE. Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. *Journal of Building Performance Simulation* 2018;11:322–32. <https://doi.org/10.1080/19401493.2017.1354919>.
45. Nutkiewicz A, Choi B, Jain RK. Exploring the influence of urban context on building energy retrofit performance: A hybrid simulation and data-driven approach. *Advances in Applied Energy* 2021;3:100038. <https://doi.org/10.1016/j.adapen.2021.100038>.
46. Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy* 2018;212:372–85. <https://doi.org/10.1016/j.apenergy.2017.12.051>.

47. Li Y, Wen Z, Cao Y, Tan Y, Sidorov D, Panasetsky D. A combined forecasting approach with model self-adjustment for renewable generations and energy loads in smart community. *Energy* 2017;129:216–27. <https://doi.org/10.1016/j.energy.2017.04.032>.
48. Liu D, Chen Q. Prediction of building lighting energy consumption based on support vector regression. 2013 9th Asian Control Conference (ASCC), IEEE; 2013, p. 1–5. <https://doi.org/10.1109/ASCC.2013.6606376>.
49. Fernandez I, Borges CE, Penya YK. Efficient building load forecasting. *Etfa2011* 2011:1–8. <https://doi.org/10.1109/ETFA.2011.6059103>.
50. Bogomolov A, Lepri B, Larcher R, Antonelli F, Pianesi F, Pentland A. Energy consumption prediction using people dynamics derived from cellular network data. *EPJ Data Science* 2016;5. <https://doi.org/10.1140/epjds/s13688-016-0075-3>.
51. Li G, Tian W, Zhang H, Fu X. A novel method of creating machine learning-based time series meta-models for building energy analysis. *Energy and Buildings* 2023;281:112752. <https://doi.org/10.1016/j.enbuild.2022.112752>.
52. Yong S-G, Kim J, Cho J, Koo J. Meta-models for building energy loads at an arbitrary location. *Journal of Building Engineering* 2019;25:100823. <https://doi.org/10.1016/j.jobee.2019.100823>.
53. Vazquez-Canteli J, Demir AD, Brown J, Nagy Z. Deep neural networks as surrogate models for urban energy simulations. *J Phys: Conf Ser* 2019;1343:012002. <https://doi.org/10.1088/1742-6596/1343/1/012002>.
54. Thrampoulidis E, Mavromatidis G, Lucchi A, Orehounig K. A machine learning-based surrogate model to approximate optimal building retrofit solutions. *Applied Energy* 2021;281:116024. <https://doi.org/10.1016/j.apenergy.2020.116024>.
55. Zhang H, Feng H, Hewage K, Arashpour M. Artificial Neural Network for Predicting Building Energy Performance: A Surrogate Energy Retrofits Decision Support Framework. *Buildings* 2022;12:829. <https://doi.org/10.3390/buildings12060829>.
56. Tardioli G, Narayan A, Kerrigan R, Oates M, O'Donnell J, Finn DP. A methodology for calibration of building energy models at district scale using clustering and surrogate techniques. *Energy and Buildings* 2020;226:110309. <https://doi.org/10.1016/j.enbuild.2020.110309>.
57. Nagpal S, Mueller C, Aijazi A, Reinhart CF. A methodology for auto-calibrating urban building energy models using surrogate modeling techniques. *Journal of Building Performance Simulation* 2019;12:1–16. <https://doi.org/10.1080/19401493.2018.1457722>.
58. Araujo G, Santos L, Leitão A, Gomes R. AD-Based Surrogate Models for Simulation and Optimization of Large Urban Areas, Sydney, Australia: 2022, p. 689–98. <https://doi.org/10.52842/conf.caadria.2022.2.689>.
59. Just Health Action, Duwamish River Cleanup Coalition. Duwamish Valley Cumulative Health Impacts Analysis. Seattle, Washington, USA: Just Health Action; 2013.
60. Duwamish Valley Program. Seattle Office of Sustainability & Environment (n.d. — page updated periodically). https://www.seattle.gov/environment/climate-change/climate-justice/duwamish-valley-program?utm_source=chatgpt.com.
61. Littell, J.S., M. McGuire Elsner, L.C. Whitely Binder, and A.K. Snover (eds). 2009. The Washington Climate Change Impacts Assessment: Evaluating Washington's Future in a Changing Climate - Executive Summary. In *The Washington Climate Change Impacts Assessment: Evaluating Washington's Future in a Changing Climate*, Climate Impacts Group, University of Washington, Seattle, Washington. n.d.
62. Jackson J, Yost M, Karr C, Fitzpatrick C, Lamb B, Chung S, et al. Public health impacts of climate change in Washington State: projected mortality risks due to heat events and air pollution 2009.
63. SHIVAv7.0.pdf n.d.
64. Clean Buildings Performance Standard (CBPS). Washington State Department of Commerce 2024. <https://www.commerce.wa.gov/cbps/> (accessed March 16, 2025).
65. Building Emissions Performance Standard - Environment | seattle.gov n.d. <https://www.seattle.gov/environment/climate-change/buildings-and-energy/building-emissions-performance-standard> (accessed March 16, 2025).
66. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput* 2011;21:137–46. <https://doi.org/10.1007/s11222-009-9153-8>.

67. Galton F. Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland* 1886;15:246. <https://doi.org/10.2307/2841583>.
68. Terry Themeau, Beth Atkinson. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13. <https://CRAN.R-project.org/package=rpart> 2018.
69. Liaw A, Wiener M. Classification and Regression by randomForest. *R News* 2002;2:5.
70. Brandon Greenwell, Bradley Boehmke, Jay Cunningham, GBM Developers. gbm: Generalized Boosted Regression Models. R package version 2.1.5. 2019.
71. Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. 2001;29:44.
72. Zou J, Han Y, So S-S. Overview of Artificial Neural Networks. In: Livingstone DJ, editor. *Artificial Neural Networks: Methods and Applications*, Totowa, NJ: Humana Press; 2009, p. 14–22. https://doi.org/10.1007/978-1-60327-101-1_2.
73. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Forth Edition. New York: Springer; 2002.
74. Gay DM. Algorithm 611: Subroutines for Unconstrained Minimization Using a Model/Trust-Region Approach. *ACM Trans Math Softw* 1983;9:503–24. <https://doi.org/10.1145/356056.356066>.
75. Kuhan A. Building Predictive Models in R Using the caret Package n.d.
76. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Soft* 2008;28. <https://doi.org/10.18637/jss.v028.i05>.
77. Gevrey M, Dimopoulos I, Lek S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling* 2003;160:249–64. [https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0).
78. Fischer A. How to determine the unique contributions of input-variables to the nonlinear regression function of a multilayer perceptron. *Ecological Modelling* 2015;309–310:60–3. <https://doi.org/10.1016/j.ecolmodel.2015.04.015>.
79. Molnar C. *Interpretable machine learning: a guide for making black box models explainable*. 3rd edition. Munich, Germany: Christoph Molnar; 2025.
80. Persily AK. Field measurement of ventilation rates. *Indoor Air* 2016;26:97–111. <https://doi.org/10.1111/ina.12193>.
81. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). *ASHRAE Handbook – Fundamentals (2017). Fundamentals (part of the annual ASHRAE Handbook series)*. Atlanta, GA: ASHRAE; 2017.
82. ENERGY STAR. Air Sealing, ENERGY STAR; 2005.
83. Air Sealing Your Home. EnergyGov n.d. <https://www.energy.gov/energysaver/air-sealing-your-home> (accessed March 14, 2025).
84. Electric Resistance Heating. Energy.gov, U.S Department of Energy (DOE); n.d.
85. Heat Pump Systems, U.S Department of Energy (DOE); n.d.
86. U.S. Environmental Protection Agency (EPA) O. *Introduction to Indoor Air Quality*, U.S. Environmental Protection Agency (EPA); 2014.
87. Low indoor temperatures and insulation. WHO Housing and Health Guidelines, World Health Organization; 2018.
88. Heat - Overview: Working in Outdoor and Indoor Heat Environments | Occupational Safety and Health Administration, U.S. Department of Labor, Occupational Safety and Health Administration; n.d.
89. International Energy Agency. *Sustainable, Affordable Cooling Can Save Tens of Thousands of Lives Each Year*. Paris, France: International Energy Agency; 2023.
90. Satish U, Mendell MJ, Shekhar K, Hotchi T, Sullivan D, Streufert S, et al. Is CO₂ an Indoor Pollutant? Direct Effects of Low-to-Moderate CO₂ Concentrations on Human Decision-Making Performance. *Environ Health Perspect* 2012;120:1671–7. <https://doi.org/10.1289/ehp.1104789>.
91. Chen C-M, Mielck A, Fahlbusch B, Bischof W, Herbarth O, Borte M, et al. Social factors, allergen, endotoxin, and dust mass in mattress. *Indoor Air* 2007;17:384–93. <https://doi.org/10.1111/j.1600-0668.2007.00486.x>.
92. Mujan I, Anđelković AS, Munćan V, Kljajić M, Ružić D. Influence of indoor environmental quality on human health and productivity - A review. *Journal of Cleaner Production* 2019;217:646–57. <https://doi.org/10.1016/j.jclepro.2019.01.307>.

93. O'Neill ZD, Li Y, Cheng HC, Zhou X, Taylor ST. Energy savings and ventilation performance from CO₂-based demand controlled ventilation: Simulation results from ASHRAE RP-1747 (ASHRAE RP-1747). *Science and Technology for the Built Environment* 2020;26:257–81. <https://doi.org/10.1080/23744731.2019.1620575>.
94. Merema B, Delwati M, Sourbron M, Breesch H. Demand controlled ventilation (DCV) in school and office buildings: Lessons learnt from case studies. *Energy and Buildings* 2018;172:349–60. <https://doi.org/10.1016/j.enbuild.2018.04.065>.
95. Mo Y, Wang C, Kassem MA, Wang D, Chen Z. Optimizing Window Configurations for Energy-Efficient Buildings with Aluminum Alloy Frames and Helium-Filled Insulating Glazing. *Sustainability* 2024;16:6522. <https://doi.org/10.3390/su16156522>.
96. Kralj A, Drev M, Žnidaršič M, Černe B, Hafner J, Jelle BP. Investigations of 6-pane glazing: Properties and possibilities. *Energy and Buildings* 2019;190:61–8. <https://doi.org/10.1016/j.enbuild.2019.02.033>.
97. Carmody J, Haglund K. *Measure Guideline: Energy-Efficient Window Performance and Selection*. U.S. Department of Energy, Energy Efficiency & Renewable Energy; 2012.
98. Abbasabadi N, Ashayeri M. Occupant-Driven Urban Building Energy Efficiency via Ambient Intelligence. In: Abbasabadi N, Ashayeri M, editors. *Artificial Intelligence in Performance-Driven Design*. 1st ed., Wiley; 2024, p. 187–209. <https://doi.org/10.1002/9781394172092.ch9>.
99. Abbasabadi N, Ashayeri M. From Tweets to Energy Trends (TwEn): An exploratory framework for machine learning-based forecasting of urban-scale energy behavior leveraging social media data. *Energy and Buildings* 2024;317:114440. <https://doi.org/10.1016/j.enbuild.2024.114440>.
100. Abbasabadi N, Ashayeri M. From tweets to energy trends (TwEn2): social sensing–informed urban building energy modeling. *Front Energy Res* 2025;13:1688348. <https://doi.org/10.3389/fenrg.2025.1688348>.
101. Ashayeri M, Abbasabadi N. Unraveling energy justice in NYC urban buildings through social media sentiment analysis and transformer deep learning. *Energy and Buildings* 2024;306:113914. <https://doi.org/10.1016/j.enbuild.2024.113914>.
102. Ashayeri M, Piri S, Abbasabadi N. Exploring U.S. Occupant Perception Toward Indoor Air Quality Via Social Media and NLP Analysis. *J Environ Sci Public Health* 2024;08. <https://doi.org/10.26502/jesph.96120205>.
103. Ashayeri M. Decoding Global Indoor Health Perception on Social Media Through NLP and Transformer Deep Learning. In: Abbasabadi N, Ashayeri M, editors. *Artificial Intelligence in Performance-Driven Design*. 1st ed., Wiley; 2024, p. 159–85. <https://doi.org/10.1002/9781394172092.ch8>.
104. Worthy A, Ashayeri M, Abbasabadi N. Leveraging earth observational data products and machine learning to enhance urban building energy modeling (UBEM) with microclimate effects. *Sustainable Cities and Society* 2025;130:106544. <https://doi.org/10.1016/j.scs.2025.106544>.
105. Ardon-Dryer K, Dryer Y, Williams JN, Moghimi N. Measurements of PM_{2.5} with PurpleAir under atmospheric conditions. *Atmos Meas Tech* 2020;13:5441–58. <https://doi.org/10.5194/amt-13-5441-2020>.
106. Tryner J, L'Orange C, Mehaffy J, Miller-Lionberg D, Hofstetter JC, Wilson A, et al. Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmospheric Environment* 2020;220:117067. <https://doi.org/10.1016/j.atmosenv.2019.117067>.

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.