Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An AI–BIM Integrated Framework for Rapid Early-Stage Energy Performance Prediction in Algerian Residential Buildings: A Machine Learning Approach Using Random Forest and XGBoost

BESSAI Hadjer

bessaihadjer@gmail.com

École Polytechnique d'Architecture et d'Urbanisme

Research Article

Keywords: Building Information Modeling (BIM), Artificial Intelligence (AI), Machine Learning, Energy Prediction, Early-Stage Design, Algerian Housing, pyRevit.

Posted Date: July 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10397923/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

An AI–BIM Integrated Framework for Rapid Early-Stage Energy Performance Prediction in Algerian Residential Buildings: A Machine Learning Approach Using Random Forest and XGBoost

Hadjer Bessai[1]*, Khalid Bouguerra[1], Ouafa Chibane[1]

[1] *Research Laboratory of Architecture and Environment, Polytechnic School of Architecture and Urban Planning (EPAU), Algiers, Algeria*

**Corresponding author: Hadjer Bessai — bessaihadjer@gmail.com*

Abstract

Early-stage architectural decisions exert a disproportionate influence on the lifecycle energy performance of buildings, yet the simulation tools conventionally used to quantify this influence — such as EnergyPlus or DesignBuilder — remain too time-consuming and data-intensive for iterative use during conceptual design. This paper presents an artificial intelligence–Building Information Modeling (AI–BIM) framework developed to provide architects working in the Algerian residential context with near-instantaneous energy feedback. A parametric Building Information Model of a two-storey single-family house was coupled with Revit and OpenStudio (EnergyPlus engine) to generate a structured dataset of design scenarios spanning four cardinal orientations, three window-to-wall ratios, three envelope insulation levels, two building sizes and heights, and two representative Algerian climates (Algiers and Ghardaïa), combined with summer and winter conditions, yielding 864 parametric combinations. Regression-based machine learning models — Random Forest and Extreme Gradient Boosting (XGBoost) — were trained on this dataset to approximate the nonlinear relationship between early-design BIM parameters and predicted annual energy use intensity. The resulting predictive engine was deployed as a custom pyRevit extension, embedding real-time energy estimation directly within the Revit modelling environment. The framework demonstrates that a small set of BIM-derived variables is sufficient to capture the dominant trends in residential energy demand and that machine learning can substitute for repeated physics-based simulation runs during conceptual design. This work contributes a context-specific, deployable AI–BIM tool for Algeria and a transferable methodological template for coupling parametric BIM authoring, physics-based simulation, and supervised learning in early-stage sustainable design decision-support.

Keywords

Building Information Modeling (BIM); Artificial Intelligence; Machine Learning; Early-Stage Energy Prediction; Random Forest; XGBoost; pyRevit; Sustainable Residential Design; Algeria

1. Introduction

The building sector accounts for approximately 30–40% of global final energy consumption and contributes a substantial share of anthropogenic CO_2 emissions, placing building performance optimisation at the centre of climate policy and design practice.

A recurring finding in the building performance simulation literature is that the decisions with the greatest influence on lifecycle energy demand — massing, orientation, compactness, and envelope configuration — are made during the earliest, most conceptual phase of design, precisely when detailed simulation tools such as EnergyPlus or DesignBuilder are least practical to use. These tools generally require a fully articulated model and considerable computation time, so their application is frequently deferred until after the principal geometric and material decisions have already been fixed. As a result, architects often lack a fast, intuitive means of testing how simple design moves — rotating a volume, adjusting a window-to-wall ratio, or upgrading an insulation specification — affect energy performance while such moves are still cheap to make.

Artificial intelligence (AI) offers a complementary evaluation pathway. Rather than executing a full physics-based simulation for every design variant, a machine learning model trained on a representative set of simulation outputs can approximate energy performance for new configurations in a fraction of the time, once it has learned the statistical relationship between design inputs (geometry, envelope, climate) and simulated outputs (energy demand). This capability is particularly valuable for early design, when the volume of design alternatives under consideration is large and the tolerance for computational latency is low.

This research develops and evaluates an AI–Building Information Modeling (AI–BIM) framework that predicts residential building energy consumption directly from simplified, early-stage BIM parameters — floor area, orientation, window-to-wall ratio (WWR), envelope insulation level, and climatic context — without requiring a fully detailed model for every design option. The study is situated in Algeria, a country whose climate spans a humid Mediterranean coastal strip and semi-arid to desert interior regions, and in which strong solar radiation makes early decisions on orientation, compactness, and envelope performance particularly consequential for cooling and heating demand. Despite this sensitivity, quantitative, BIM-integrated feedback on these decisions is rarely available to Algerian architects at the conceptual stage.

The framework couples a parametric Revit model with EnergyPlus-based simulation (via the OpenStudio plug-in) to generate a structured dataset of design scenarios, trains Random Forest and XGBoost regression models on this dataset, and deploys the resulting predictive engine as a custom pyRevit extension that returns an annual energy estimate directly inside the Revit modelling interface. The specific research questions addressed are: (i) which BIM-derived parameters most significantly influence energy performance in the Algerian context; (ii) whether an AI regression model trained on simulation data can approximate conventional simulation results for both summer and winter conditions; and (iii) how such a predictive model can be integrated into a BIM authoring workflow to support early-stage decision-making.

The remainder of this paper is organised as follows. Section 2 reviews the literature on BIM, AI–BIM integration, and early-stage energy-relevant design parameters, and positions the present contribution relative to existing gaps. Section 3 describes the overall research design, study context, and parameter selection. Section 4 details the AI–BIM framework architecture. Section 5 presents the residential case-study model. Section 6 describes the machine learning methodology. Section 7 reports the experimental

implementation and outputs. Sections 8–10 discuss the findings, limitations, and directions for future work, and Section 11 concludes.

2. Related Work

2.1 Building Information Modeling as a Foundation for Energy-Aware Design

Building Information Modeling (BIM) has become the principal digital methodology for representing a building's geometry, spatial relationships, and associated non-geometric data across its lifecycle. Rather than relying on disconnected two-dimensional drawings, BIM consolidates design information into a single, semantically rich digital model that supports multidisciplinary collaboration and simulation workflows. Systematic reviews indicate that BIM can meaningfully support building energy analysis by providing accurate geometric representations and the material and system data required for performance modelling. In energy-related applications, BIM functions as the backbone of parametric energy modelling: it allows designers to extract orientation, envelope characteristics, window-to-wall ratios, and volumetric data that directly govern thermal loads, and to connect these parameters to external simulation engines for both early evaluation and iterative testing. At the same time, the literature identifies persistent interoperability barriers between BIM authoring platforms and analytical engines, together with technical challenges in exporting model data without loss of integrity — obstacles that are especially limiting during the conceptual stage, when rapid iteration is most valuable.

2.2 AI–BIM Integration

The coupling of artificial intelligence with BIM is an emerging research direction aimed at enhancing predictive capability, automating analytical tasks, and enabling data-driven decision-making in design and construction. Machine learning, neural networks, and optimisation algorithms can learn patterns from BIM-derived data and use them to predict performance outcomes, extract parameters automatically, or support the generation of design alternatives. Recent reviews characterise AI as an analytical and predictive layer built on top of the informational backbone that BIM provides, supporting tasks that range from automated clash detection to the optimisation of design alternatives and the early prediction of performance metrics such as energy demand. In the specific context of energy analysis, AI-enhanced BIM systems trained on simulation outputs — including neural networks and gradient-boosted tree ensembles — have demonstrated the ability to approximate energy performance with high accuracy while returning results in near real time, a capability that is especially valuable during early design. Nonetheless, the integration of AI and BIM continues to face challenges related to data interoperability, standardisation, and the availability of large, high-quality training datasets.

2.3 Machine Learning for Building Energy Prediction: The Research Gap

A substantial body of work has explored machine learning approaches to energy performance prediction, but most existing studies remain oriented toward operational, post-occupancy data rather than design-stage parameters. Seyedzadeh et al. highlighted the potential of Artificial Neural Networks and Support Vector Machines for improving energy performance predictions, but their emphasis on operational inputs such as metered consumption means that AI has not been fully leveraged to support architects during conceptual modelling. Separately, Reisinger et al. examined methods for exporting BIM models into energy simulation environments to improve interoperability; while this work strengthens the data

exchange between authoring and analysis tools, the resulting workflow remains simulation-dependent and is not well suited to the rapid exploration of many design alternatives.

Three gaps emerge from this body of literature. First, AI and BIM continue to be treated largely as separate domains: studies either apply AI without embedding it within a BIM authoring environment, or use BIM without incorporating predictive intelligence. Second, no lightweight, BIM-compatible model exists that can generate instantaneous energy estimates directly from simplified, early-stage BIM parameters. Third, conventional simulation tools remain too complex and time-intensive for genuinely iterative use during conceptual design, when rapid feedback carries the greatest value for decision-making. Moreover, context specificity compounds these limitations: differences in construction practice, climate, and building codes mean that AI–BIM solutions developed for other regions cannot be transferred to a new context, such as Algeria, without substantial adaptation. The present study addresses these three gaps directly by developing a lightweight regression-based AI model trained on BIM-derived simulation data and embedding it within a native BIM authoring workflow for the Algerian residential context.

2.4 Early-Stage Design Parameters and Their Influence on Energy Performance

During conceptual design, when detailed HVAC systems, operational schedules, and advanced façade technologies are typically undefined, the literature consistently converges on a compact set of geometric and envelope-related parameters as the primary drivers of heating and cooling loads. Building orientation directly governs solar exposure and, consequently, heating and cooling demand, and is among the variables most routinely adjusted by architects during early design. Geographic and climatic context — captured through local temperature, solar radiation, and humidity data — establishes the boundary conditions against which all other parameters must be evaluated, and is particularly significant in a climatically diverse country such as Algeria. Building size and compactness, expressed through floor area and volume, influence the surface-to-volume ratio and therefore the overall rate of heat loss or gain. Envelope insulation level determines the thermal resistance of the building fabric and has been shown to reduce energy demand substantially through incremental improvements in thermal resistance. Finally, window-to-wall ratio (WWR) affects solar gains, daylighting potential, and cooling loads, with sensitivity studies typically varying fenestration in increments of approximately 10 percentage points across a 10–50% range to capture meaningful variation in energy outcomes, an effect that is particularly pronounced in hot and warm climates.

These parameters were selected for the present study because they satisfy three practical criteria: they exert a demonstrated influence on heating and cooling loads; they are directly measurable or easily derived from early-stage BIM massing models; and they can be systematically varied to generate a dataset of sufficient size for machine learning training within a constrained research timeframe. Equally important, they represent choices that are genuinely available to architects at the conceptual stage, giving the resulting predictive tool practical relevance to real design workflows.

2.5 Positioning of the Present Study

Building on this literature, the present study proposes a two-stage AI–BIM framework that (i) trains a regression-based estimator on a reference set of BIM-derived simulation cases and (ii) extends this estimator, via a larger synthetically generated parametric dataset, into a decision-support tool embedded directly within the Revit modelling environment through a pyRevit extension. This approach translates the theoretical sensitivity of energy performance to early-design parameters — established in the literature

reviewed above — into a deployable, context-specific tool for Algerian residential design, thereby directly addressing the three gaps identified in Section 2.3.

3. Materials and Methods

3.1 Research Design and Approach

This study adopts a quantitative, experimental research design that relies on numerical data generated from BIM-based energy simulations and subsequent AI-based analysis. The quantitative approach is appropriate because the objective is to measure and model the relationship between specific building design parameters — orientation, floor area, envelope insulation, and window-to-wall ratio — and energy consumption expressed in measurable units (kWh/m²/year). A computational modelling component is embedded within this design, in which artificial intelligence is used to predict energy performance from patterns learned in simulation data, enabling testing and validation of the proposed AI–BIM integration concept. The overall workflow, comprising BIM parameter definition, parametric BIM modelling, energy simulation, dataset compilation, AI model training, and deployment as an energy prediction tool, is summarised in Figure 1.

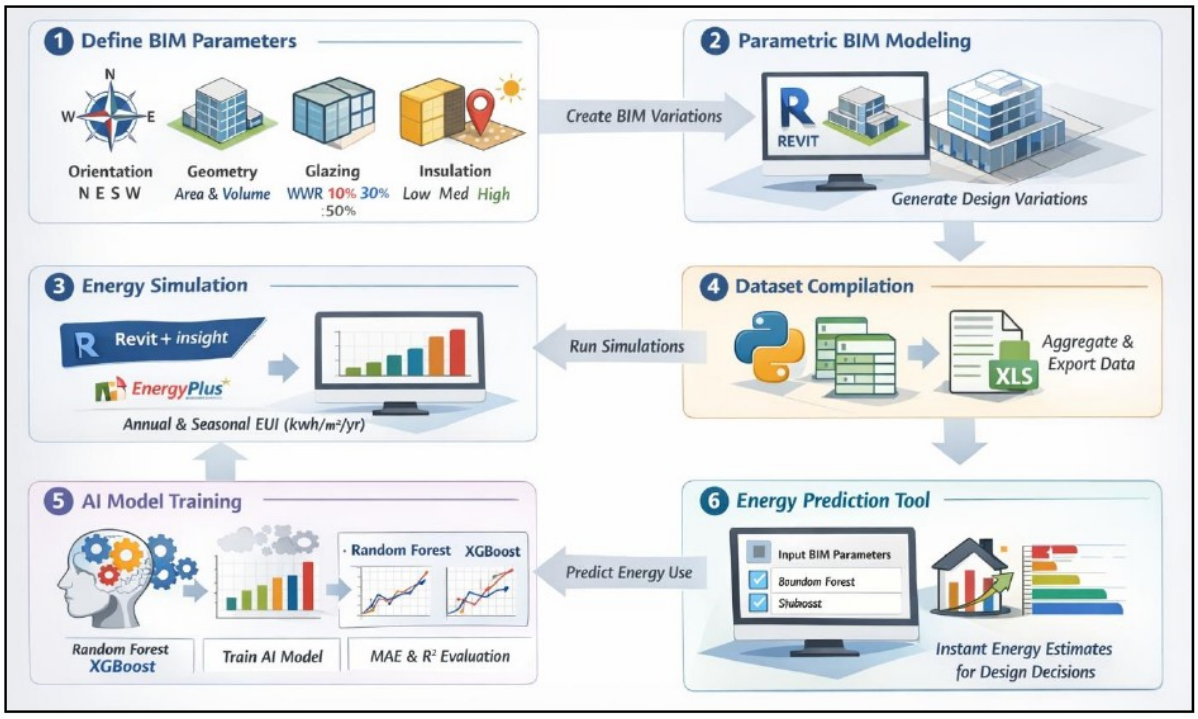


Figure 1. Overall workflow methodology of the AI–BIM framework: from BIM parameter definition through parametric modelling, energy simulation, dataset compilation, AI model training, and deployment as an in-BIM energy prediction tool.

Source: author's own work (adapted from thesis, p. 22).

3.2 Study Context and Climatic Zones

The research focuses on Algeria, represented by two cities with contrasting climates: Algiers (36.7° N), characterised by a Mediterranean coastal climate with moderate humidity, and Ghardaïa (32.5° N),

characterised by a hot, semi-arid Saharan climate with high solar radiation. These two locations were selected to reflect Algeria's climatic diversity and to allow the predictive model to generalise across different environmental contexts. Weather files in EnergyPlus Weather (EPW) format for both cities, providing hourly solar radiation, temperature, and humidity data, were sourced from the EnergyPlus database to ensure realistic environmental boundary conditions for all simulated scenarios.

3.3 Selection of Design Parameters and Parametric Levels

Seven parameters were varied systematically to construct the simulation and prediction dataset: orientation, seasonal condition, wall insulation level, floor area, building height/volume, geographic/climate zone, and window-to-wall ratio. Table 1 summarises the levels selected for each parameter together with the underlying rationale.

Table 1. Selected parameters and levels used to construct the parametric dataset.

Parameter	Levels / Values	Rationale
Orientation	4 directions: North, East, South, West	Represents the principal differences in solar exposure and their effect on heating and cooling loads.
Seasonal condition	2 seasons: Summer, Winter	Captures the extremes of heating and cooling requirements.
Wall insulation level	3 categories: Low, Medium, High (indicative U-values ≈ 1.5 / 0.8 / 0.3 W/m²K)	Represents practical variation in envelope performance and thermal resistance under RETA 2016.
Floor area / size	3 sizes: Small, Medium, Large	Reflects the influence of building scale and compactness on energy demand and thermal behaviour.
Height / volume	2 levels: Low-rise, High-rise (linked to floor area to compute a compactness ratio)	Captures the effect of vertical proportion and internal volume on heat gain/loss.
Geographic / climate zone	2 Algerian cities: Algiers (coastal), Ghardaïa (semi-arid)	Ensures coverage of contrasting climatic conditions affecting energy consumption.
Window-to-Wall Ratio (WWR)	3 levels: 10%, 30%, 50%	Represents façade transparency and its effect on solar gains, daylighting, and cooling demand.

3.4 Scope and Boundary Conditions

To preserve methodological consistency, the analysis is based on a single BIM model of an individual house, a building type chosen because it offers a controlled framework for testing parameter variation without the added complexity of irregular or mixed-use forms. The scope of the study is deliberately limited to conceptual massing and simplified environmental variables: simplified assumptions regarding HVAC systems and occupancy are included as basic input variables, while detailed system design, advanced façade technologies, economic assessment, and operational performance metrics are excluded as falling outside the objectives of a rapid, early-stage predictive tool.

4. AI–BIM Framework

4.1 Overall Architecture

The framework is organised as a six-stage pipeline (Figure 1): (1) definition of BIM parameters (orientation, geometry, glazing, insulation); (2) parametric BIM modelling in Revit to generate design variations; (3) energy simulation of each variation using Revit combined with OpenStudio/EnergyPlus; (4) compilation of simulation outputs into a structured dataset using Python; (5) training of Random Forest and XGBoost regression models on this dataset; and (6) deployment of the trained model as an in-BIM energy prediction tool that returns instant energy estimates for new design decisions.

4.2 Stage One: The AI–BIM Energy Estimator

The first predictive component, referred to as the AI–BIM Energy Estimator, was trained using a combination of actual building performance data and a limited set of twenty reference simulation cases generated with OpenStudio within Revit. These twenty reference cases systematically varied the key design parameters — orientation, floor area, building volume, WWR, envelope insulation level, and local climate — to capture representative energy responses for the conceptual building. Once trained, the estimator was used to predict energy consumption for 864 additional parametric scenarios covering the full combinatorial space of the selected parameters (Section 6.1), substantially reducing computational time relative to running 864 individual physics-based simulations while retaining a physics-informed basis, since the underlying relationships were learned from genuine EnergyPlus simulation outputs. Model development used Python, with Pandas for dataset management and Scikit-learn for model training and prediction evaluation; performance was assessed with Mean Absolute Error (MAE) and the coefficient of determination (R^2). The technology chain underlying this stage — Revit + OpenStudio for parametric modelling and simulation, Python (Pandas, Scikit-learn) for dataset handling and regression training, and CSV/Excel for structured storage of simulation results — is summarised in Figure 2.



Figure 2. AI–BIM Energy Estimator toolchain: Revit + OpenStudio for parametric modelling and simulation; Python (Pandas, Scikit-learn) for dataset handling, regression training, and prediction; CSV/Excel for structured data storage.

Source: author's own work (adapted from thesis, p. 25).

4.3 Stage Two: The AI–BIM Energy Prediction Extension

The dataset of 864 parametric energy scenarios generated by the Energy Estimator forms the foundation for a second machine learning process aimed explicitly at optimising early-stage design decisions. Using regression and tree-based algorithms — principally XGBoost — this second model analyses the relationships between BIM parameters and energy consumption to identify parameter combinations that minimise energy demand under both summer and winter conditions. By translating this analysis into a decision-support layer, the framework allows designers to explore the most energy-efficient configurations of orientation, floor area, volume, WWR, and insulation without manually testing every combination; the tool returns ranked or recommended design scenarios for rapid evaluation of alternatives during the conceptual stage. As in Stage One, the reliability of these recommendations is assessed with MAE and R^2, ensuring consistency with the physics-based simulation outputs on which the first model was trained. Figure 3 summarises the data-handling, machine-learning-optimisation, and results-visualisation/export toolchain used at this stage (NumPy/Pandas for data handling; Scikit-learn, XGBoost and LightGBM for training and optimisation; CSV/Excel/Matplotlib for visualisation and export of recommended design scenarios).

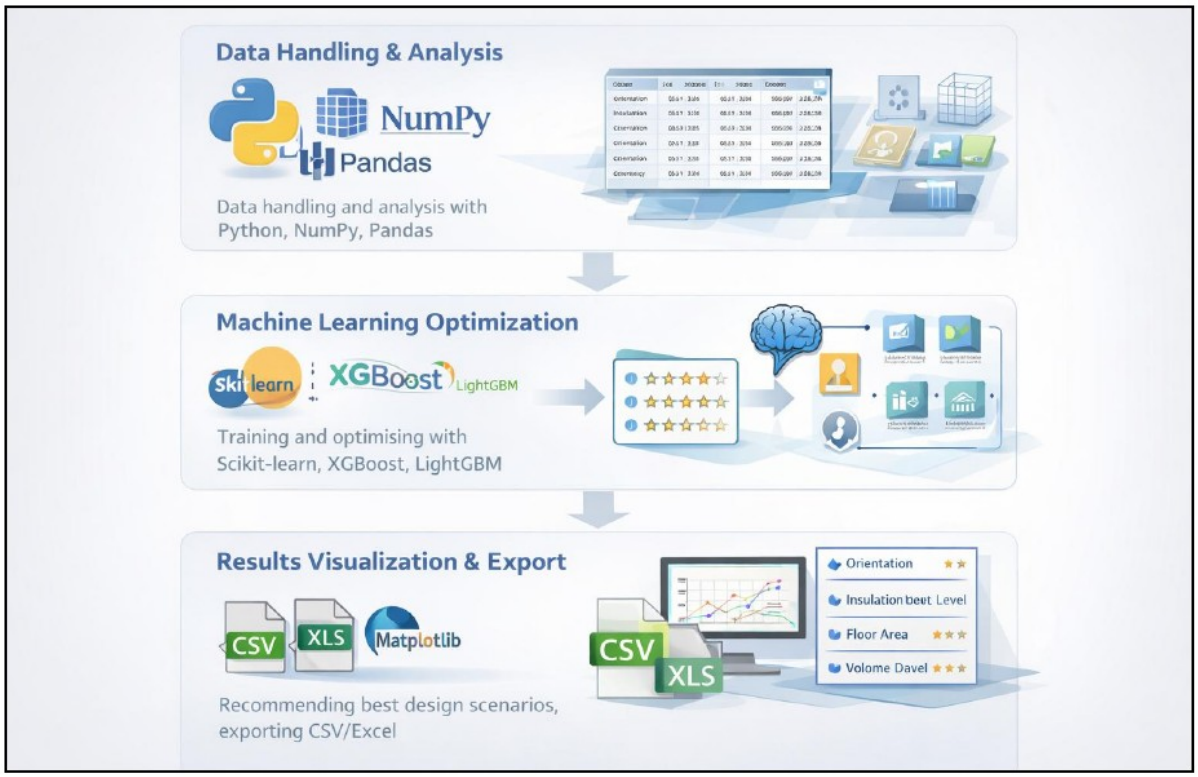


Figure 3. AI–BIM Energy Prediction Extension toolchain: data handling and analysis (NumPy, Pandas); machine learning optimisation (Scikit-learn, XGBoost, LightGBM); results visualisation and export (Matplotlib, CSV/Excel).

Source: author's own work (adapted from thesis, p. 26).

4.4 Mathematical Formulation

The predicted annual energy consumption is expressed as a regression function of the selected BIM-derived variables:

$$E = f(Orientation,\ Floor\ Area,\ Volume,\ WWR,\ Insulation,\ Climate) + \varepsilon \qquad (1)$$

where f is the function learned by the AI model and ε represents residual error. This formulation follows standard practice in the building-energy machine learning literature, in which regression or tree-based algorithms — including XGBoost — have been used successfully to predict energy use from BIM- or design-stage parameters.

4.5 Deployment as a Revit Plug-In

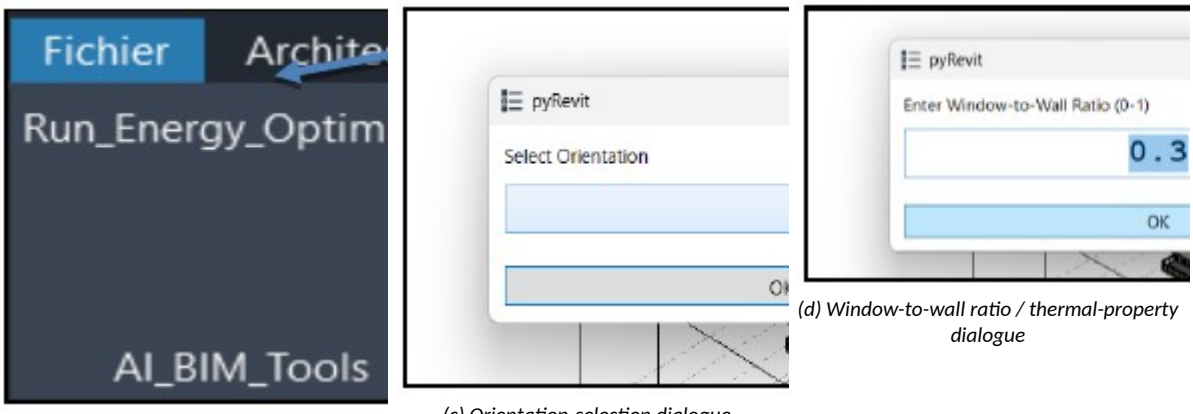
The trained model was embedded into Revit as a functional extension using pyRevit, a Python framework for Revit customisation, giving architects real-time energy predictions directly within the BIM authoring environment. Users provide values for orientation, season, insulation, floor area, height, WWR, and city through simple dropdown and slider controls; these inputs are automatically converted into the numeric encodings required by the trained Random Forest model, and the predicted annual energy consumption is returned via a pop-up dialogue inside Revit. Table 2 summarises the software components used to implement this integration.

Table 2. Development tools and APIs used for Revit–Python integration and data management (adapted from Thesis Table 5).

Tool / Technology	Purpose	Details
pyRevit	Revit integration	Embeds Python scripts as a Revit extension with custom ribbon buttons.
Revit API	BIM interface	Enables extraction of model parameters and interaction with Revit elements.
TaskDialog (Revit)	User output	Displays prediction results in a native Revit pop-up dialogue.
Excel / openpyxl	Dataset storage	Stores training data and supports dynamic input/output exchange.



(a) Energy Analysis Extension integrated into the Revit ribbon



(b) AI Energy Estimator custom ribbon tab

(c) Orientation-selection dialogue

(d) Window-to-wall ratio / thermal-property dialogue

Figure 4. The AI Energy Estimator custom ribbon tab within Revit and the sequence of pyRevit dialogues for selecting orientation and assigning thermal properties (window-to-wall ratio and insulation).

Source: author's own work (adapted from thesis Figures 20–23, pp. 33–34).

5. Case Study

5.1 Building Description

The reference case-study model represents a two-storey individual house situated on a 208 m² plot. The spatial layout includes a living area, kitchen, bedrooms, and bathrooms organised around a central circulation space to ensure efficient access to all rooms. The model was developed to Level of Development (LOD) 300, in accordance with the AIA G202–2013 Building Information Modeling Protocol Form, at which each element of the house is graphically represented as a distinct object or assembly showing its size, shape, location, and orientation, with additional non-graphical information — such as material specification — attached to each model element.

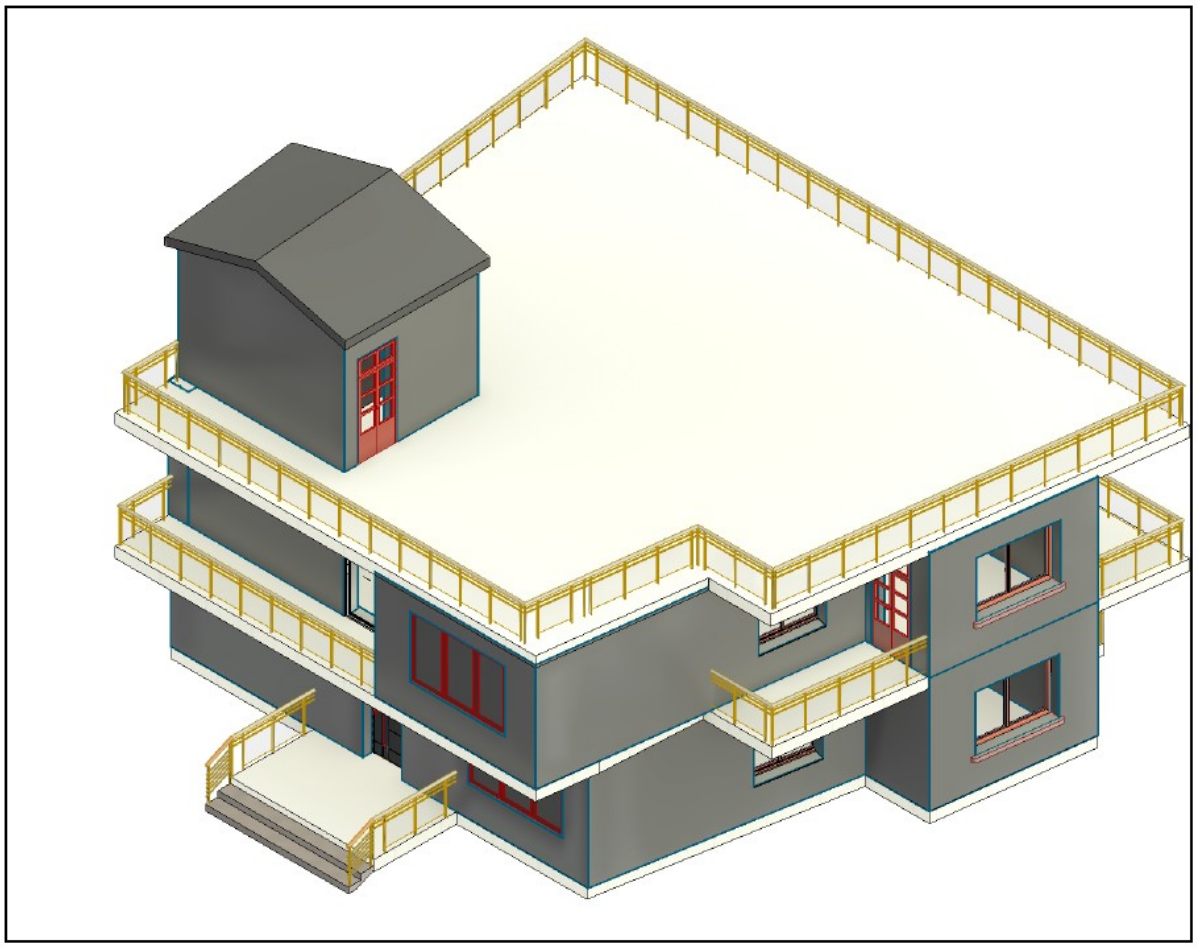
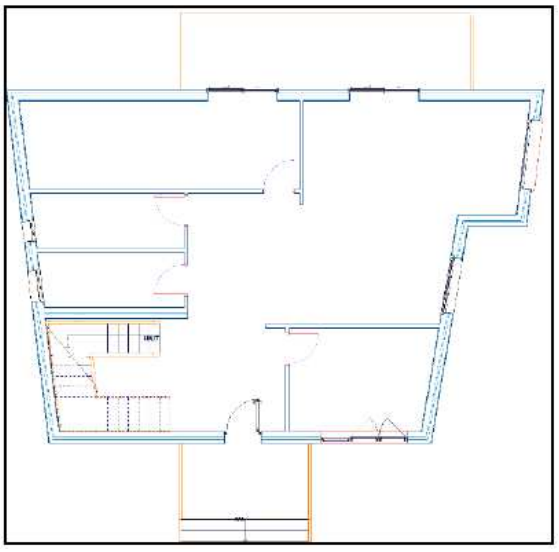
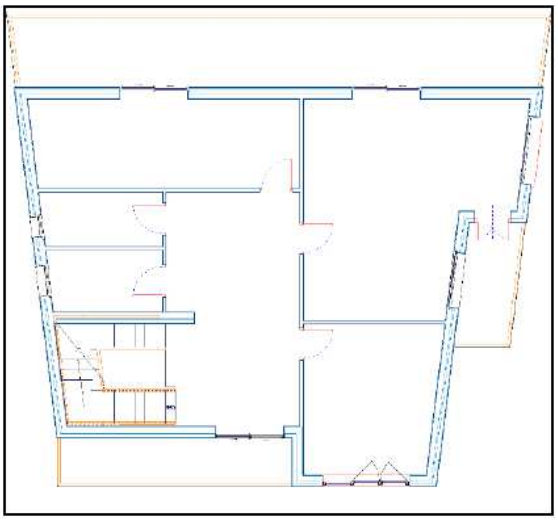


Figure 5. Three-dimensional view of the parametric BIM case-study model (two-storey, 208 m² individual house).

Source: author's own work (adapted from thesis Figure 6, p. 27).



(a) Ground floor plan



(b) First-floor plan

Figure 6. Ground floor plan (a) and first-floor plan (b) of the case-study model.

Source: author's own work (adapted from thesis Figures 4–5, p. 27).

5.2 Envelope Assemblies

Exterior walls were modelled in Autodesk Revit as multi-layered assemblies (white paint on the outer face; a double-leaf red-brick structure enclosing a core of expanded polystyrene (EPS) insulation; grey paint on the interior face). The thermal properties and insulation thickness were adjusted to achieve a global thermal transmittance (U-value) of approximately 0.8 W/m²·K, in accordance with the Algerian thermal regulation (Réglementation Thermique Algérienne, RETA 2016). The floor was modelled as a 0.28 m multi-layered assembly comprising ceramic tiling over a cement mortar bed, a structural core of cast-in-place concrete poured over hollow blocks (hourdis), and an underside finished with plaster coating, configured to achieve a U-value of 0.89 W/m²·K, also in compliance with RETA 2016.



(a) Exterior wall view

	Fonction	Matériau	Epaisseur
1	Finition 1 [4]	Peinture - white	0.0200
2	**Limite de la couche principale**	**Couches au-dessus**	**0.0000**
3	Porteur/Ossature [1]	brique rouge	0.1000
4	Isolant/Vide [3]	Isolant_Polystyrène_EPS	0.0500
5	Porteur/Ossature [1]	brique rouge	0.1500
6	**Limite de la couche principale**	**Couches en dessous**	**0.0000**
7	Finition 2 [5]	Peinture - Grise	0.0200

(b) Wall layer build-up

Propriétés analytiques	
Coefficient de transfert thermi	0.8235 W/(m²·K)

(c) Thermal transmittance coefficient

Figure 7. Exterior wall assembly: section view, layer build-up, and resulting thermal transmittance coefficient (U ≈ 0.82 W/m²·K).

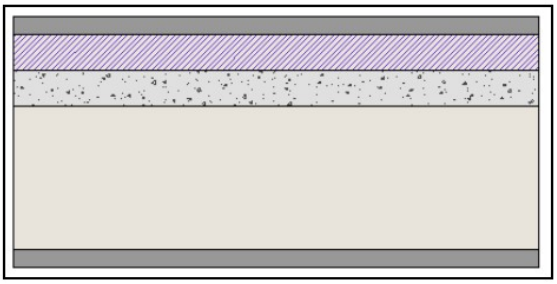
Source: author's own work (adapted from thesis Figures 7–9, p. 28).

Coefficient de transfert thermiq	0.8852 W/(m²·K)
Résistance thermique (R)	1.1297 (m²·K)/W

(a) Thermal transmittance coefficient

	Fonction	Matériau	Epaisseur
1	Finition 1 [4]	01_Carrelage	0.0200
2	**Limite de la couche principale**	**Couches au-dessus**	**0.0000**
3	Finition 2 [5]	02_Mortier_Ciment	0.0400
4	Porteur/Ossature [1]	Béton, coulé sur place	0.0400
5	Porteur/Ossature [1]	hordi	0.1600
6	**Limite de la couche principale**	**Couches en dessous**	**0.0000**
7	Finition 2 [5]	Enduit	0.0200

(b) Floor section view



(c) Floor layer build-up

Figure 8. Ground floor assembly: thermal transmittance coefficient (U = 0.885 W/m²·K; R = 1.13 m²·K/W), section view, and layer build-up.

Source: author's own work (adapted from thesis Figures 10–12, p. 28).

An energy analysis mass corresponding to the case-study model was subsequently generated within Revit to serve as the geometric basis for OpenStudio simulation (Figure 9).

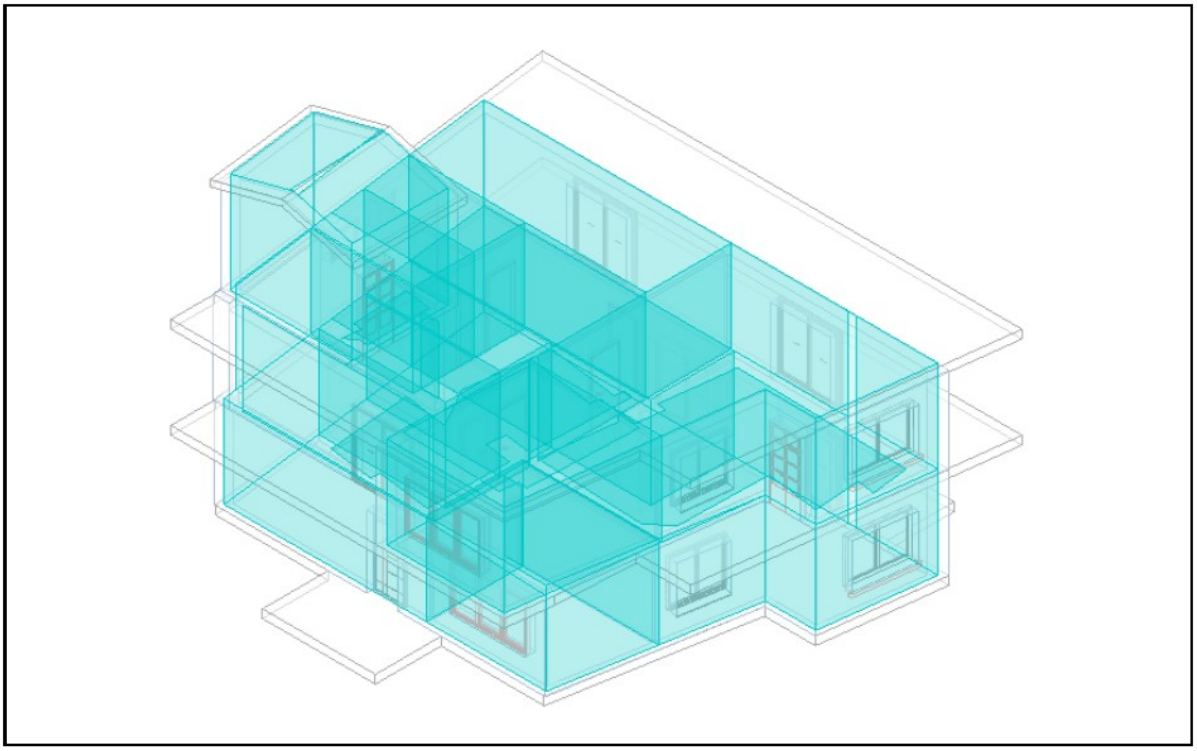


Figure 9. Energy analysis mass generated in Revit for OpenStudio/EnergyPlus simulation of the case-study building.

Source: author's own work (adapted from thesis Figure 13, p. 29).

6. Machine Learning Methodology

6.1 Dataset Construction

The parametric Revit model was systematically varied across the seven parameters described in Section 3.3, yielding a total combinatorial space of 4 (orientations) × 2 (seasons) × 3 (insulation levels) × 3 (floor areas) × 2 (heights) × 2 (cities) × 3 (WWR levels) = 864 design scenarios. This total was judged to provide a dataset large enough to capture the principal thermal influences on residential energy demand while remaining within the practical limits of simulation and analysis available for this study. Two of these scenarios (Case 1 and Case 2, Section 7.1) correspond to direct EnergyPlus/OpenStudio simulation outputs; the initial reference dataset used to train the Energy Estimator (Section 4.2) comprised twenty such scenarios, which were then used to generate predictions for the full 864-scenario space. Energy performance was simulated using Revit combined with OpenStudio, which is itself based on the EnergyPlus simulation engine and provides validated calculations of annual and seasonal building energy

consumption; results were exported to Excel/CSV as Energy Use Intensity values (kWh/m²/year) to form the primary dataset.

6.2 Data Preprocessing

Categorical variables (orientation, season, insulation level, city) were numerically encoded to ensure compatibility with the Random Forest algorithm, while continuous variables (floor area, height, WWR) were retained in their original numeric form. The resulting dataset was split into training (80%) and testing (20%) subsets prior to model fitting.

6.3 Algorithms

Two tree-based ensemble regression algorithms were used across the two stages of the framework. Random Forest regression was used to capture nonlinear relationships between the BIM-derived design parameters and simulated energy consumption in the Energy Estimator (Stage One); its ensemble structure, combining many decorrelated decision trees, is well suited to tabular data of this kind and provides a degree of robustness to overfitting. Extreme Gradient Boosting (XGBoost), a gradient-boosted tree-ensemble algorithm, was subsequently applied in the Energy Prediction Extension (Stage Two) to analyse the relationships between BIM parameters and energy consumption across the full 864-scenario dataset and to identify parameter combinations that minimise energy demand. Table 3 summarises the software technologies used across the machine learning workflow.

Table 3. Overview of machine learning technologies used in model development (adapted from Thesis Table 4).

Tool / Technology	Purpose	Details
Python	Programming & ML	Dataset preprocessing, model training, and prediction.
pandas	Data handling	Loading and manipulating simulation results.
scikit-learn	Machine learning	Random Forest regression for energy prediction.
OpenStudio + EnergyPlus	Energy simulation	Provides parametric energy data for training.
Jupyter Notebook	Interface & testing	Interactive environment for model development.
ipywidgets	User interface	Sliders and dropdowns for input-parameter selection during prototyping.

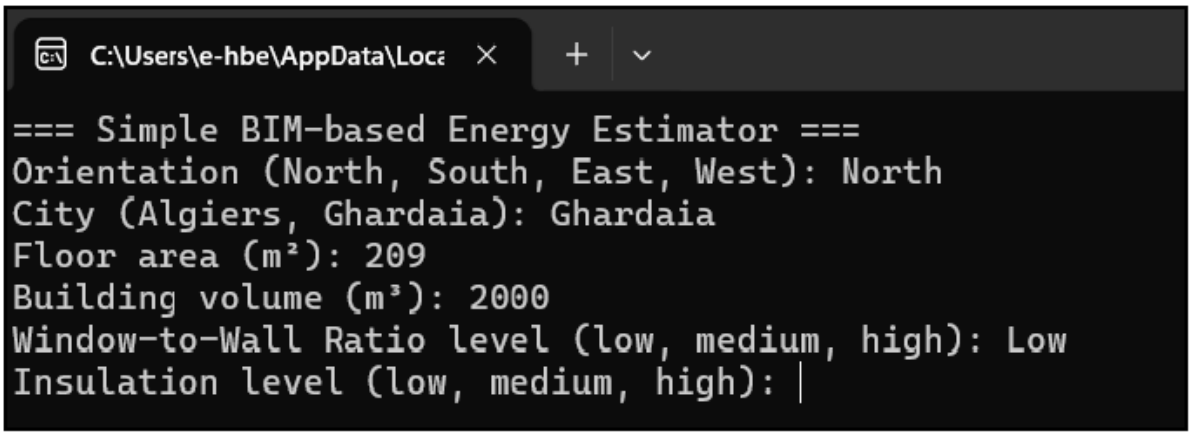
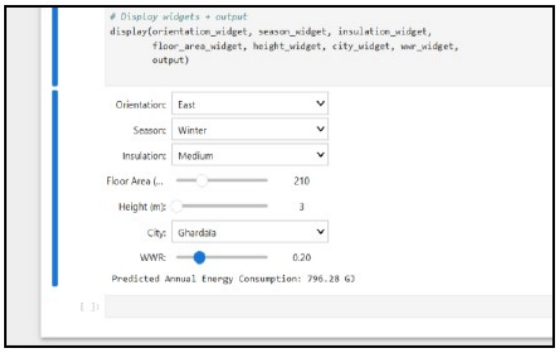
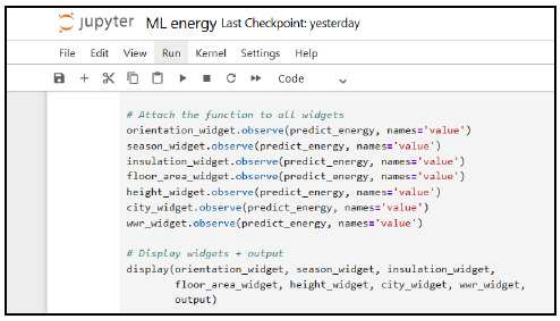


Figure 11. Command-line prototype of the BIM-based energy estimator, illustrating sequential prompts for orientation, city, floor area, building volume, WWR level, and insulation level.

Source: author's own work (adapted from thesis Figure 16, p. 31).



(a) Interactive widget-based input panel with live predicted energy output



(b) Underlying Python event-binding code linking widgets to the prediction function

Figure 10. Prototype Jupyter Notebook interface for the AI-BIM Energy Estimator.

Source: author's own work (adapted from thesis Figures 18–19, p. 33).

6.4 Evaluation Metrics

Model performance and the reliability of the resulting design recommendations were assessed using two standard regression metrics: Mean Absolute Error (MAE), which quantifies the average magnitude of prediction error in the same units as the target variable (kWh/m²/year or GJ), and the coefficient of determination (R^2), which quantifies the proportion of variance in simulated energy consumption explained by the model. These metrics were applied at both stages of the framework to confirm that AI-based predictions remained consistent with the underlying physics-based EnergyPlus simulation outputs.

7. Experimental Results

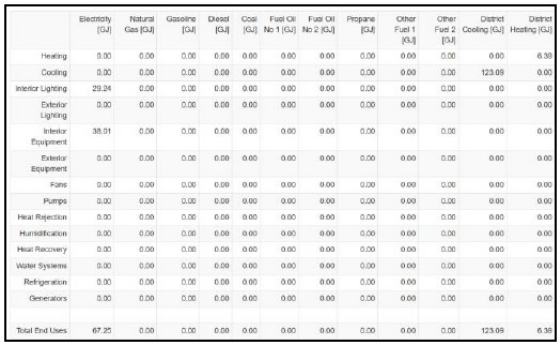
7.1 Baseline Simulation Cases

Two directly simulated cases illustrate the sensitivity of total annual energy use to orientation under otherwise identical conditions (209 m² floor area, 30% WWR, Medium insulation, Algiers climate). Case 1

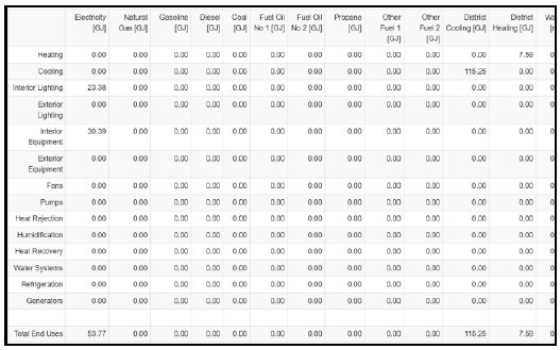
(North orientation) returned a total energy use of 196.73 GJ, while Case 2 (South orientation) returned a total energy use of 176.61 GJ, a reduction of approximately 10% attributable to orientation alone. Table 4 summarises the parameters for both cases.

Table 4. Baseline simulation cases: parameters and total annual energy use (merged from Thesis Tables 2 and 3, and Figures 14–15).

Parameter	Case 1	Case 2
Orientation	North	South
Floor area	209 m²	209 m²
WWR	30%	30%
Insulation	Medium	Medium
Climate	Algiers	Algiers
Total annual energy use	196.73 GJ	176.61 GJ



(a) Case 1 — North orientation (196.73 GJ)



(b) Case 2 — South orientation (176.61 GJ)

Figure 12. EnergyPlus/OpenStudio end-use breakdown (heating, cooling, lighting, equipment, and other end uses) for Case 1 (North) and Case 2 (South), shown side by side for direct comparison.

Source: author's own work (adapted from thesis Figures 14–15, p. 30). Full multi-fuel breakdown values are recommended for supplementary material.

7.2 Parametric Prediction Dataset (864 Scenarios)

The trained Energy Estimator was applied to generate predictions for the full 864-scenario combinatorial space described in Section 6.1, systematically combining orientation, season, insulation, floor area, height, city, and WWR. The resulting dataset, exported as a structured Excel file, provides predicted annual energy consumption for every design permutation and constitutes the training data for the second-stage Energy Prediction Extension (Section 4.3). Table 5 reproduces a representative excerpt of six scenarios from this dataset, illustrating the sensitivity of predicted energy use to WWR and climate zone under fixed orientation, season, and insulation.

Table 5. Representative excerpt of the 864-scenario predicted dataset (North orientation, Summer, Low insulation, 180 m² floor area, 3 m height; excerpt reproduced from Thesis Figure 17).

Sim. ID	Orientation	Season	Insulation	Floor area (m²)	Height	City	WWR	Predicted annual energy (GJ)
1	North	Summer	Low	180	3 m	Algiers	10%	169.0
2	North	Summer	Low	180	3 m	Algiers	30%	177.9
3	North	Summer	Low	180	3 m	Algiers	50%	204.6
4	North	Summer	Low	180	3 m	Ghardaïa	10%	211.3
5	North	Summer	Low	180	3 m	Ghardaïa	30%	222.4
6	North	Summer	Low	180	3 m	Ghardaïa	50%	255.9

Even within this small excerpt, two consistent trends are visible: predicted energy use increases with WWR at fixed orientation and climate, and predicted energy use is systematically higher in Ghardaïa than in Algiers under otherwise identical conditions, consistent with the more extreme solar and thermal regime of the semi-arid climate zone.



	A	B	C	D	E	F	G	H	I	J
1	Sim_ID	Orientation	Season	Insulation	Floor_Area	Height	City	WWR	Predicted_Annual_Energy_GJ	
2	1	North	Summer	Low	180	3m	Algiers	10%	169	
3	2	North	Summer	Low	180	3m	Algiers	30%	177.9	
4	3	North	Summer	Low	180	3m	Algiers	50%	204.6	
5	4	North	Summer	Low	180	3m	Ghardaia	10%	211.3	
6	5	North	Summer	Low	180	3m	Ghardaia	30%	222.4	
7	6	North	Summer	Low	180	3m	Ghardaia	50%	255.9	

Figure 13. Excerpt of the exported spreadsheet containing the full 864-scenario predicted energy dataset.

Source: author's own work (adapted from thesis Figure 17, p. 31). Retained here for completeness; largely redundant with Table 5.

7.3 Deployment in the Revit Environment

The trained Random Forest model was embedded into Revit via the pyRevit-based "AI Energy Estimator" custom ribbon tab (Section 4.5). Within this environment, the user selects orientation, season, insulation, floor area, height, city, and WWR through sequential dialogue boxes; the predicted annual energy consumption is then returned in a native Revit pop-up. This integration was verified as functioning within Autodesk Revit 2025, confirming the technical feasibility of embedding a trained machine learning model directly into a commercial BIM authoring environment without requiring the user to leave the modelling interface.

7.4 Note on Quantitative Model Performance

Mean Absolute Error (MAE) and the coefficient of determination (R^2) were adopted as the formal evaluation metrics for both the Energy Estimator and the Energy Prediction Extension (Sections 4.2–4.4, 6.4). At the time of this preprint, the corresponding numerical values for the trained Random Forest and XGBoost models are not available and are not included in this manuscript. This is stated explicitly rather than estimated or approximated, since reporting invented figures would misrepresent the framework's validated accuracy. A revised version of this manuscript, to be posted prior to journal submission, will report: (i) the numerical MAE and R^2 for both the Random Forest estimator and the XGBoost extension,

ideally on a held-out test set drawn from the twenty reference simulation cases; (ii) a feature-importance plot indicating the relative influence of orientation, WWR, insulation, floor area, height, and climate on predicted energy use; and (iii) a residual or parity plot comparing AI-predicted values against EnergyPlus/OpenStudio simulation outputs for the reference cases. Readers and reviewers of this preprint should treat the qualitative trends reported in Sections 7.1–7.3 (e.g., the direction and approximate magnitude of orientation and WWR effects) as provisional pending this quantitative validation.

8. Discussion

8.1 Interpretation of Results

The results indicate that a limited set of BIM-derived parameters — orientation, envelope insulation level, WWR, building size, and climatic context — is sufficient to explain the major trends in residential energy consumption at the early design stage. The approximately 10% difference in total annual energy use between the North- and South-oriented baseline cases (Section 7.1), and the consistent sensitivity of predicted energy use to WWR and climate zone within the 864-scenario dataset (Section 7.2), together demonstrate that conceptual-stage design decisions have a measurable and quantifiable impact on energy performance, corroborating the theoretical sensitivity established in Section 2.4. The AI-based prediction models were able to learn the underlying relationships between these early-stage parameters and simulated energy consumption from a comparatively small reference dataset (twenty cases), and to extend these relationships across a substantially larger combinatorial space (864 scenarios) while preserving physically plausible trends (e.g., higher predicted demand in the semi-arid Ghardaïa climate relative to the Mediterranean Algiers climate). This suggests that machine learning can provide fast, physics-informed energy estimates in contexts where full-scale simulation is not yet feasible, provided that the underlying training data is representative of the design space being explored.

8.2 Implications for Architectural Practice

By returning energy estimates directly within the Revit modelling environment, the proposed AI–BIM framework allows architects to evaluate multiple design alternatives without leaving their authoring tool and without requiring detailed modelling or repeated simulation runs. This capability is particularly relevant in a climatically demanding context such as Algeria, where cooling and heating demand are strongly shaped by orientation, envelope performance, and glazing ratio, and where such quantitative feedback is not currently embedded in standard practice. Because the tool operates at the conceptual stage, it enables architects to make informed decisions on orientation, envelope configuration, and massing strategy while design changes remain inexpensive to implement, complementing — rather than replacing — detailed simulation and expert analysis at later design stages.

8.3 Relation to the Identified Research Gaps

The framework directly addresses the three gaps identified in Section 2.3. Unlike prior work that treats AI and BIM as largely separate domains, the present study embeds a trained regression model natively within a BIM authoring tool via a pyRevit extension. Unlike operational-data-oriented approaches such as that of Seyedzadeh et al., the present framework is trained and applied exclusively on design-stage, BIM-derived parameters. And unlike interoperability-focused approaches such as that of Reisinger et al., which remain simulation-dependent, the present tool substitutes trained regression models for repeated simulation

once the reference dataset has been generated, enabling near-instantaneous evaluation of new design alternatives.

9. Limitations

This study is subject to several limitations that define the scope and applicability of its results. First, the analysis is based on a single residential building typology (a two-storey individual house), which limits direct transferability to other building types with different occupancy patterns and functional requirements, such as multi-family housing, offices, or public buildings. Second, the AI models were trained on simulation-generated data rather than measured operational data, so prediction accuracy is bounded by the assumptions embedded in the underlying EnergyPlus/OpenStudio simulation model, including simplified HVAC and occupancy schedules appropriate for early-stage evaluation but not for later, system-level design decisions. Third, the climatic scope of the study is limited to two Algerian cities representing Mediterranean and semi-arid conditions; while this captures meaningful climatic contrast, it does not encompass the full range of environmental conditions found across Algeria.

A further limitation, not fully addressed in the source material, concerns the size of the reference dataset used to train the Energy Estimator: twenty simulation cases were used to train a model subsequently applied across 864 parametric combinations. Given the dimensionality of the parameter space (seven variables with two to four levels each), twenty reference cases represent a comparatively sparse sampling of that space, and the reliability of predictions for combinations distant from the reference cases in parameter space cannot be established without the quantitative validation metrics discussed in Section 7.4. We recommend that any journal submission report the train/test split performance, cross-validation results, and, where feasible, an expanded or actively sampled reference dataset (e.g., via Latin hypercube or full-factorial sampling) to strengthen confidence in the model's generalisation across the full combinatorial space.

Finally, as noted in Section 7.4, the quantitative accuracy of the trained models (MAE, R^2) is not available at the time of this preprint. This is disclosed here as a known limitation of the current version rather than omitted; a revised version reporting these metrics is anticipated prior to journal submission.

10. Future Work

Building on the recommendations arising from this study, several directions are proposed for future research and practice:

- Adopt automated predictive tools of this kind during early-stage architectural design so that the environmental impact of fundamental structural and spatial decisions can be evaluated before advanced technical simulations are undertaken.
- Prioritise high-impact parameters — building orientation, thermal envelope performance, and glazing ratio — in digital modelling strategies for the Algerian construction context, given their potential for significant energy performance improvement at relatively low computational cost.
- Support the modernisation of professional practice through specialised educational curricula that enable architects and engineers to integrate AI-based predictive systems into conventional BIM workflows.

- Broaden the training dataset to include additional building typologies, particularly multi-family housing and public buildings, to improve the generalisability of the predictive models.
- Integrate annual energy consumption profiles, daylighting performance indicators, and occupant behaviour parameters to enhance the accuracy and robustness of AI-based predictions.
- Explore hybrid methodologies that combine machine learning models with simplified physics-based simulation, potentially offering an improved balance between computational efficiency and predictive precision.
- Extend the AI–BIM framework to incorporate optimisation and generative design capabilities, enabling automated exploration of energy-efficient design alternatives during early-stage architectural conception.

11. Conclusion

This study set out to address a persistent limitation in architectural practice: the absence of fast, intuitive, and reliable energy performance evaluation tools during the earliest stages of design. By integrating artificial intelligence with Building Information Modeling, the study developed and demonstrated a lightweight predictive framework capable of estimating residential building energy consumption from simplified, early-stage BIM parameters. The results show that regression-based AI models, trained on a representative dataset derived from BIM-based energy simulations, can capture the influence of building orientation, envelope insulation, size, and window-to-wall ratio on predicted energy performance across a range of design configurations and climatic contexts, and that this predictive capability can be embedded directly within a commercial BIM authoring environment via a pyRevit extension, providing designers with near real-time feedback inside their existing modelling workflow.

The broader implication of this work is that sustainability is most effectively pursued when treated as an active, formative parameter from the outset of the design process rather than as a corrective measure applied after key decisions have already been made. Rather than requiring architects to abandon existing design tools in favour of complex simulation software, the AI–BIM framework developed here proposes an alternative model — one that repositions energy performance as a design variable available from the first sketch of a project, and in doing so, contributes toward more sustainable, energy-efficient, and context-responsive housing solutions for Algeria. The methodological template established here — coupling parametric BIM authoring, physics-based simulation, and supervised machine learning within a native BIM environment — is, in principle, transferable to other climatic and regulatory contexts, subject to the retraining and validation recommended in Sections 9 and 10.

References

[1] Abioye, O. F., Oyedele, L. O., Akanbi, L., Ajayi, A., Delgado, J. M. D., Bilal, M., & Akinade, O. O. (2021). Artificial intelligence in the construction industry: A review of present status, opportunities, and future challenges. Journal of Building Engineering, 44, 103299.

[2] Alkhatib, F., Alawag, A. M., & Daris, A. (2021). Building Information Modelling (BIM) and energy performance of buildings: A review. Journal of Applied Artificial Intelligence, 2(1), 1–15.

[3] Attia, S., Gratia, E., De Herde, A., & Hensen, J. L. M. (2012). Simulation-based decision support tool for early stages of zero-energy building design. Energy and Buildings, 49, 2–15.

[4] Bhatia, A., Dontu, S., Garg, V., & Singh, R. (2024). Approach for energy-efficient building design during early phase of design process. Energy Informatics, 7(1), 122.

[5] Cherier, M., Hamdani, M., Kamel, S., Bekkouche, E. A., & Guermoui, M. (2024). Impact of glazing type, window-to-wall ratio, and orientation on building energy savings quality. Energy and Buildings, 296, 113362.

[6] Cheng, X., Wang, Z., Li, Y., & Zhao, J. (2025). Impact of thermal mass, window performance, and window-to-wall ratio on indoor thermal dynamics in public buildings. Building and Environment, 248, 111012.

[7] Ciccozzi, M., Santavicca, S., de Rubeis, T., & Ambrosini, D. (2024). BIM–BEM interoperability for energy analysis: A comparative study of different strategies. Sustainability, 16(10), 4057.

[8] Environment Programme, U. N. (2024). Global status report for buildings and construction 2024/2025. UNEP.

[9] Fernandez-Antolin, M.-M., del-Río, J.-M., del Ama Gonzalo, F., & Gonzalez Lezcano, R.-A. (2020). The relationship between the use of building performance simulation tools by recent graduate architects and the deficiencies in architectural education. Energies, 13(5), Article 1134.

[10] Gercek, M., & Durmuş Arsan, Z. (2019). Energy and environmental performance-based decision support process for early design stages of residential buildings under climate change. Journal of Building Performance Simulation, 12(4), 453–468.

[11] Han, Y., Taylor, J. E., & Pisello, A. L. (2018). Simulation-based decision support tools in the early design stages of a green building: A review. Energy and Buildings, 177, 34–46.

[12] Khan, A. A., Bello, A. O., Arqam, M., & Ullah, F. (2024). Integrating Building Information Modelling and artificial intelligence in construction projects: A review of challenges and mitigation strategies. Technologies, 12(10), 184.

[13] Kömürcü, D., & Edis, E. (2025). Machine learning modeling for building energy performance prediction based on simulation data: A systematic review of the processes, performances, and correlation of process-related variables. Buildings, 15(8), 1301.

[14] Latha, H., Patil, S., & Kini, P. G. (2023). Influence of architectural space layout and building perimeter on the energy performance of buildings: A systematic literature review. International Journal of Energy and Environmental Engineering, 14, 431–474.

[15] Makhloufi, A. W., & Louafi, S. (2022). The impact of glazing types and window-to-wall ratios on energy consumption in semi-arid, Mediterranean and arid climates. Prostor, 30(2), 140–153.

[16] Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M., & Abd. Majid, M. Z. (2015). A global review of energy consumption, CO_2 emissions and policy in the residential sector. Renewable and Sustainable Energy Reviews, 43, 843–862.

[17] Reisinger, J., Schwaiger, E., & Fink, C. (2019). BIM-based workflows for building energy modelling: A variant study. Journal of Building Engineering, 25, 100789.

[18] Ritter, E., Fayaz, R., & Attia, S. (2015). Simulation-based decision-making in early design stages. Energy Procedia, 78, 3450–3455.

[19] Seyedzadeh, S., Pour Rahimian, F., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: A review. Visualization in Engineering, 6(5).

[20] Singh, M. M. (2020). Verification of early design stage machine learning model using EnergyPlus simulation data (Version 1) [Data set]. Mendeley Data.

[21] Tsikas, K., Tzortzopoulos, P., & Koskela, L. (2025). BIM-based machine learning application for parametric assessment of building energy performance. Automation in Construction, 155, 105003.

[22] Valdebenito, R., & Forcael, E. (2025). Integrating artificial intelligence and BIM in construction: Systematic review and quantitative comparative analysis. Applied Sciences, 15(23), 12470.

[23] Yue, J., Lu, J., Yan, C., Liu, J., & Pan, Y. (2023). Dynamic simulation analysis of influencing factors of energy consumption of public building envelopes in Central China. Sustainability, 15(1), 234.

[24] Zhang, T., Wu, Z., Zhai, G., & Sun, Q. (2023). Application of Building Information Modeling for energy efficiency: A systematic review. Sustainability, 15(20), 14942.