ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.2224

Mathematical Modeling of Energy Consumption in Residential HVAC Systems as an Input Variable for an Intelligent Control System

Juan José Laínez Bolaños*, Douglas Antonio Plaza Guingla†, Carolina Alexandra Casal Quintero‡, and Juan José Laínez Casal§

*†*Escuela Superior Politécnica del Litoral, Facultad de Ingeniería en Electricidad y Computación, Guayaquil, Ecuador*
‡*Escuela Superior Politécnica del Litoral, Facultad de Ingeniería Mecánica y Ciencias de la Producción, Guayaquil, Ecuador*
§*University of Central Florida, Department of Mechanical Engineering, Florida, United States of America*
E-mail: jjlainez@espol.edu.ec , douplaza@espol.edu.ec , caralcas@espol.edu.ec , juan.lainez@ucf.edu

Abstract*– Modern heating, ventilation, and air conditioning (HVAC) systems account for between 40% and up to 60% of the total energy consumption in commercial and residential buildings, which presents significant opportunities for optimization through intelligent control strategies. Traditional HVAC control systems operate reactively, adjusting only to the set temperature without considering energy consumption patterns or optimization. This study develops a comprehensive mathematical model of HVAC energy consumption that serves as a dynamic input variable for advanced control systems, enabling predictive climate control with energy-awareness. The proposed methodology integrates artificial neural networks for hourly energy consumption prediction, considering city-specific meteorological variables, HVAC equipment parameters, indoor characteristics of the conditioned space, and real consumption data from HVAC systems across 34 brands, 70 models, 9 BTU/h capacities, and 4 compressor technologies, including On-Off, Inverter, Digital Inverter, and Dual Inverter. The model combines thermodynamic principles with machine learning techniques to predict real-time energy consumption. This research contributes to the development of intelligent and energy-efficient HVAC systems, essential for the transition toward sustainable buildings.

***Keywords:* HVAC control, energy consumption modeling, model predictive control, building energy management, thermal comfort optimization,.**

I. Introduction

Building energy consumption represents approximately 40% of global energy consumption and 36% of total CO_2 emissions. HVAC systems constitute the most significant component, accounting for 40% to 60% of energy used in buildings [1, 2]. As global energy demand continues to increase and climate change concerns intensify, optimizing HVAC system performance has become crucial for building sustainability [3].

Traditional HVAC control systems employ reactive strategies, typically using proportional-integral-derivative (PID) controllers that respond to temperature deviations from setpoint values [4, 12]. While simple and reliable, these systems present several fundamental limitations, including lack of predictive capability or inability to anticipate changes in thermal load, energy-blind operation because energy consumption and costs are not considered, inefficient response to disturbances since adjustment is reactive rather than proactive, suboptimal equipment operation with frequent on-off cycles, and inefficiency at partial load [2, 4].

Accurate prediction of future energy consumption is particularly challenging due to the inherent nonlinearity of thermodynamic processes, stochastic variability of occupancy patterns, and complex dependence on multiple meteorological variables [1, 8]. Traditional methods based on detailed physical models require extensive parameterization and significant computational power, limiting their applicability in real-time control systems [10, 13].

The study was conducted in households along the coast of Ecuador, in the cities of Guayaquil, Daule, Durán, El Triunfo, Isidro Ayora, Jipijapa, Naranjito, Nobol, Pascuales, Pedro Carbo, Petrillo, Salitre, and Tenguel. A total of 4,200 electrical energy consumption measurements and ambient temperature measurements were collected, along with equipment and site information. The measurement period was between November 16 and 28, 2023. The equipment types studied ranged from 5,000 to 60,000 BTU/h, including window and split types, with current market technologies shown in Table 1.

TABLE 1
HVAC EQUIPMENT TECHNOLOGIES MEASURED IN THIS STUDY

Technology	Compressor Operation	Energy Efficiency	Typical COP	Typical SEER
On/Off	Full speed start/stop	Low	2.5 - 3.2	9 - 13
Inverter	Variable speed	High	3.2 - 4.2	14 - 17
Digital Inverter	Precision digital variable speed	Very high	3.8 - 4.8	16 - 22
Dual Inverter	Dual-rotary variable speed	Highest	4.0 - 5.2	18 - 26

In On/Off technology, the compressor operates at constant maximum speed; upon reaching the set temperature, it turns off. When the ambient temperature rises again, it restarts at maximum power. This technology has high starting current, frequent start-stop cycles, greater temperature fluctuations resulting in lower efficiency, higher energy consumption, greater mechanical wear, and higher noise during startup [15, 18]. The advantages of this technology include lower initial cost, plain design, and easy maintenance.

Inverter technology uses a frequency converter to continuously vary the compressor speed [14, 19]. Instead of stopping, the compressor reduces speed once the target temperature is reached. This reduces start-stop cycles, improves temperature stability, reduces energy consumption by 20% to 40% compared to On-Off, produces less noise, extends compressor lifespan, but has a higher initial cost and more complex electronics [15, 18].

Digital Inverter technology uses an advanced inverter with microprocessor digital control, providing more precise modulation of compressor speed and more precise adaptation to load, more stable temperature control, better efficiency at partial load, improved power factor, and energy optimization, with lower noise levels, although the cost is slightly higher than standard inverter [14, 15].

Dual Inverter technology uses a dual rotary compressor (two compression chambers), which reduces vibration and improves load distribution, operates with variable frequency like Inverter technology, provides faster cooling response, has the highest efficiency among residential systems, extended compressor lifespan, and has the highest purchase cost on the market [15, 19].

Regional manufacturers commonly use metrics such as COP and SEER [5, 18]. The **Coefficient of Performance (COP)** is the ratio between useful energy produced (cooling or heating) and electrical energy consumed [5, 18]:

COP = Thermal Energy Transferred (kW) / Electrical Energy Consumed (kW) (1)

However, it has the following limitations: measured under specific conditions (35°C exterior), does not reflect performance at different temperatures, does not consider seasonal use, and is an instantaneous value, not annual [5].

The **Seasonal Energy Efficiency Ratio (SEER)** measures air conditioning efficiency during an entire cooling season (not just at a specific point) [5, 18]:

SEER = Total Seasonal Cooling (BTU) / Total Energy Consumed (kWh) (2)

This metric has the following limitations: it is based on specific U.S. climate conditions, not universal (different regions use different metrics), more complex to calculate, and difficult to verify independently [5, 18].

The HVAC equipment studied share common characteristics including 220 V at 60 Hz supply voltage and R-410A refrigerant. Data was obtained from manufacturer datasheets of brands such as TCL, LG, and Samsung. For national brands, manufacturer information was not available; therefore, actual consumption measures were taken.

The objectives of this study are to develop a mathematical model of HVAC system energy consumption, to validate model accuracy with actual measured data, to provide implementation guidelines for residences or other sites where commercial split equipment is used, and for this modelling to be part of the input variables for a feasible and economical control system to implement [4, 9].

II. Methodology

A. *Case Study and Population*

The study was conducted in households distributed across Guayaquil, Daule, Durán, El Triunfo, Isidro Ayora, Jipijapa, Naranjito, Nobol, Pascuales, Pedro Carbo, Petrillo, Salitre, and Tenguel; The Ecuadorian coastal region comprises all localities along the Pacific coast and has a predominantly tropical climate with warm, humid conditions and marked spatial variation: the north is more humid and rainy, while the south is drier and warmer. Temperatures range from 25 °C to 33 °C, with a wet season from December to April and a dry season from May to November. [20].

The study population consists of measurements collected from 221 air conditioning units, yielding a total of 4,201 observations. Due to limitations in material, human, technical, and financial resources, the sample was restricted to the student community of the Faculty of Industrial Engineering at the University of Guayaquil, specifically individuals occupying indoor spaces equipped with air conditioning systems. Data collection was conducted from November 16 to 28, 2023.

A survey was conducted to collect the following data: brand, HVAC equipment model, temperature, and energy consumption measurement from the electric company meter at each household with the HVAC equipment turned off, meter reading with the HVAC equipment turned on.

During a two-hour period, ambient temperature was taken every 5-minute, given a total of 24 measurements per household. A database was obtained with the following data: HVAC equipment identification, equipment brand, equipment model, date, time of each measurement, length, width, and height of the conditioned space, window area of the space, ambient temperature, and electrical consumption during the experiment.

To validate the conditions at each household with the standard ASHRAE [5], calculations were made for each site data was taken. It was determined that the installed equipment capacity is adequate, the following was considered: heat transmission through walls, ceiling, and windows; internal heat gain from people (sensible and latent heat); lighting; 3 electrical equipment; ventilation and outdoor air infiltration;

solar gains through windows and sun-oriented surfaces; and interior and exterior design conditions [5, 6]..

B. System Architecture and Variables

Measurement Instrument Design

The questionnaire was designed with six sections and twenty-four questions, following questionnaires according to ANSI/ASHRAE 55 and ISO 8996 standards [5, 6]. The sections considered in the questionnaire structure were as shown in Table 2.

TABLE 2
HVAC EQUIPMENT, METEOROLOGICAL, SITE, AND TEMPERATURE AND ENERGY CONSUMPTION MEASUREMENT VARIABLES

Section	Variable	Description
HVAC Equipment Information	ID	HVAC equipment identification
	M	HVAC equipment brand
	MO	HVAC equipment model
	ET	Equipment type
	CT	Compressor technology
	CCKW	Cooling capacity (kW)
	RIPKW	Rated input power (kW)
	COP	Coefficient of performance (COP)
	SEER	Seasonal Energy Efficiency Ratio (SEER)
	EC	Energy class
	RF	Refrigerant
	MS	Manufacturer source (PDF/URL)
	BTUI	Installed HVAC capacity (BTU/h)
Meteorological Information	C	City
	AVT	Average Temperature of City
	HRC	Relative Humidity % of City
	SR	Solar Radiation W/m²
Site Information	AN	Site width where the HVAC equipment is installed
	L	Site length where the HVAC equipment is installed
	AR	Site area where the HVAC equipment is installed
	AL	Site height where the HVAC equipment is installed
	V	Site volume where the HVAC equipment is installed
	NP	Number of occupants at the site
Measurements	F	Measurement date
	BTUD	Calculated HVAC capacity (BTU/h)
	RBTU	Difference between BTUI and BTUD
	H	Measurement time
	TA	Ambient temperature
	KWH	Energy consumption (kWh)

The methodology for getting the raw data is the following.

1. The student fills in the survey containing table 2.
2. The student enters the room with the HVAC system, where he will turn on the unit a write down the first temperature measured by a temperature sensor with a precision of ±0.1.
3. The student records the temperature in intervals of 5 minutes for two hours.
4. The student turns off the unit
5. The student reports the data.

C. Energy Consumption Prediction Model

Main Model Formulation

The complete equation of the energy consumption variable kWh as a function of the variables is:

$$kWh = \mu_y + \sigma_y \cdot W^{(5)} \cdot \varphi_4(W^{(4)} \cdot \varphi_3(W^{(3)} \cdot \varphi_2(W^{(2)} \cdot \varphi_1(W^{(1)} \cdot ((x - \mu_x)/\sigma_x) + b^{(1)}) + b^{(2)}) + b^{(3)}) + b^{(4)}) + b^{(5)} \quad (3)$$

where:

$\varphi_l(z) = \mathrm{ReLU}(\mathrm{BatchNorm}(z))$ for layers $l = 1, 2, 3, 4$

x = vector of 11 input variables

Θ = 48,257 trainable parameters

TABLE 3
NEURAL NETWORK INPUT VECTOR VARIABLES

Variable	Description	Unit
x_1 = TA	Ambient Temperature	°C
x_2 = BTUI	Installed Capacity	BTU/h
x_3 = RBTU	Difference BTUI-BTUD	BTU/h
x_4 = AVT	City Average Temperature	°C
x_5 = HRC	Relative Humidity	%
x_6 = SR	Solar Radiation	W/m²
x_7 = AR	Site Area	m²
x_8 = V	Site Volume	m³
x_9 = NP	Number of Occupants	persons
x_{10} = C	City (encoded)	-
x_{11} = CT	Compressor Technology	-

Output Variable**:** y = kWh = Energy Consumption (kWh)

Network Architecture

Layer Structure:

Input (11)

Layer 1: [11 → 256] + BatchNorm + ReLU + Dropout(0.3), 3,584 Parameters

Layer 2: [256 → 128] + BatchNorm + ReLU + Dropout(0.3), 33,152 Parameters

Layer 3: [128 → 64] + BatchNorm + ReLU + Dropout(0.2), 8,384 Parameters

Layer 4: [64 → 32] + BatchNorm + ReLU + Dropout(0.2), 2,144 Parameters

Output: [32 → 1] (Linear), 33 parameters. Total: 47,297 parameters

Complete Technical Proposal for IoT HVAC Optimization System

Given that this study shows that most equipment has On-Off compressor technology, a low-cost hardware and software solution can be implemented, with economic investment easily recoverable through energy savings. This approach aligns with recent developments in IoT-based HVAC system modelling [7,14].

Efficiency Improvement

TABLE 4
HVAC EQUIPMENT EFFICIENCY IMPROVEMENT

System	COP Improvement	SEER Improvement	Energy Savings
On-Off	18-30%	38-50%	20-35%
Inverter	23-35%	40-55%	25-40%

Hardware

The technical proposal for implementing an intelligent IoT control system that improves COP and SEER of HVAC systems includes:

a) CPU: Raspberry Pi 4 (4GB RAM)

b) Temperature and Humidity Sensors, SHT31-D: Precision: ±0.3°C, ±2% Relative Humidity, I2C Interface

c) Occupancy Sensors, Passive Infrared Sensor: HC-SR501, Movement detection 3-7m

d) Energy Monitoring (kWh), PZEM-004T, Precision ±1%

e) Current transformers: SCT-013-030. For On-Off Compressors: Smart relay with four channels Shelly 4PM. For Inverter Compressors: Modbus to 0-10V Converter

III. RESULTS

D. *Data Processing*

The data and measurements were processed in MATLAB and Python, obtaining the following results shown in the subsequent figures.

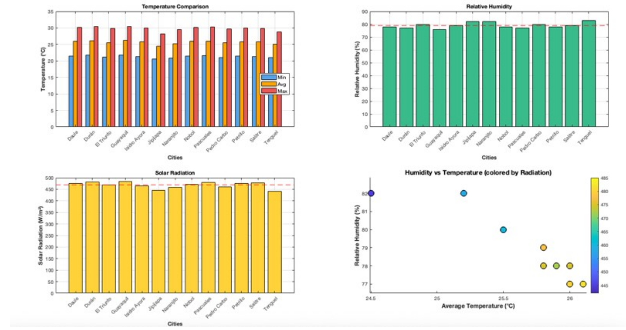

Figure 1. Averages, maximums, and minimums of temperatures (top left), relative humidity (top right), solar radiation (down left), and the relationship among these variables in the cities where the measurements were taken (down right).

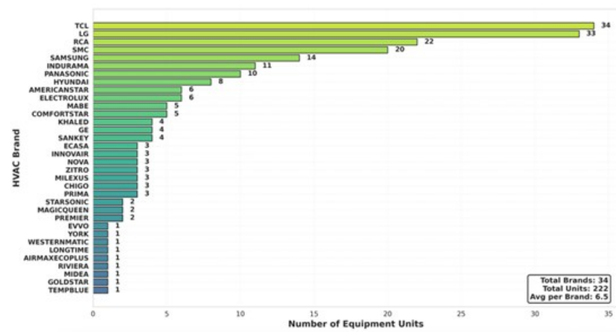

Figure 2. Horizontal bar chart showing the distribution of HVAC equipment units across different manufacturers in the dataset.

Brands are ranked by frequency in Fig.2, with bar colors transitioning through a viridis colormap from darker (fewer units) to lighter yellow green (more units). Each bar is labelled with the exact count of equipment units. A statistics box in the lower right corner provides summary information including total number of brands, total units, and average units per brand. The figure reveals that the data set contains equipment from multiple major HVAC manufacturers, with varying representation levels.

The brand distribution analysis provides critical context for understanding the dataset's composition and potential biases. The varying representation levels across brands (ranging from single units to 40+ units) suggest that the dataset includes both major market players and niche manufacturers. The concentration of certain brands may reflect either market dominance in the study region or sampling bias toward specific distributors or installation companies. The presence of well-known international brands (LG, Samsung, Panasonic, Electrolux) alongside regional manufacturers indicates a diverse HVAC market ecosystem. This diversity is beneficial for model development as it ensures that the trained neural network learns patterns generalizable across different equipment technologies rather than being biased toward specific manufacturer characteristics.

However, the uneven distribution also means that prediction accuracy may vary by brand, with better performance expected for well-represented manufacturers.

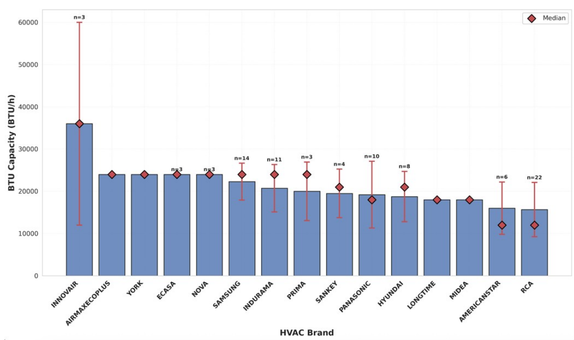

Figure 3. Average BTU capacity across the top fifteen brands ranked by capacity.

Figure 3 shows average BTU capacity across the top fifteen brands ranked by capacity. The figure uses a bar chart format where each bar represents the mean BTU capacity for a brand, with error bars indicating one standard deviation above and below the mean. Red diamond markers overlaid on the bars indicate the median capacity for each brand, providing a comparison between mean and median that reveals distribution skewness. Small annotations above each bar show the sample size (n=X) for each brand. The x-axis labels show brand names, while the y-axis displays capacity in BTU/h.

The capacity analysis reveals significant variation in typical equipment sizes across manufacturers, with mean capacities ranging from 12,000 to 24,000 BTU/h. The magnitude of standard deviation bars (often comparable to or exceeding the mean) indicates that most brands offer equipment across a wide capacity range rather than focusing on narrow segments. Discrepancies between mean and median (when diamond markers are notably higher or lower than bar centres) suggest non-normal distributions, typically right-skewed distributions where a few high-capacity units pull the mean upward. Brands positioned at the higher end of capacity ranges may specialize in commercial applications or larger residential spaces, while those at lower capacities focus on single-room or small-space applications. The capacity information is critical for energy modelling because cooling/heating capacity is a primary determinant of potential energy consumption, though actual consumption depends on utilization patterns and efficiency ratings.

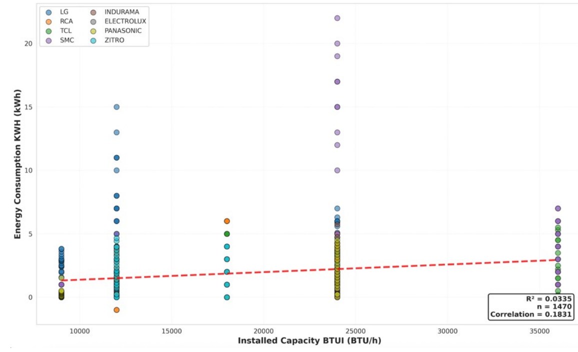

Figure 4. Scatter plot shows the relationship between installed capacity (BTUI, x-axis) and actual energy consumption (kWh, y-axis) for the top eight brands by measurement count.

Figure 4 presents a scatter plot showing the relationship between installed capacity (BTUI, x-axis) and actual energy consumption (kWh, y-axis) for the top eight brands by measurement count. Each brand is represented by a distinct colour, with data points displayed as filled circles with black edges for clarity. A red dashed line overlays the scatter plot representing a linear regression fit through all data points, with the regression equation displayed in the legend. A statistics box in the lower right corner provides the R^2 value, total sample size (n), and Pearson correlation coefficient. The legend identifies each brand by colour and includes the trend line information.

This scatter plot reveals how capacity-consumption relationships vary across manufacturers. The overall trend line (red dashed) represents the average capacity-consumption relationship across all brands, while the distribution of individual brand clusters around this line indicates whether specific manufacturers achieve better or worse efficiency than average. Brands whose points systematically fall below the trend line operate more efficiently (less consumption for given capacity), while those above the line are less efficient. The scatter magnitude within each brand cluster reflects operational variability, and tight clusters indicate consistent performance, while dispersed clouds suggest diverse usage patterns or installation conditions. The R^2 value quantifies how much of consumption variation is explained by capacity alone; low R^2 values (e.g., <0.3) indicate that capacity is insufficient to predict consumption, and other factors (usage patterns, efficiency ratings, environmental conditions) dominate. The slope of the trend line reveals the marginal energy impact of each additional BTU of capacity—steeper slopes mean greater energy penalties for larger equipment.

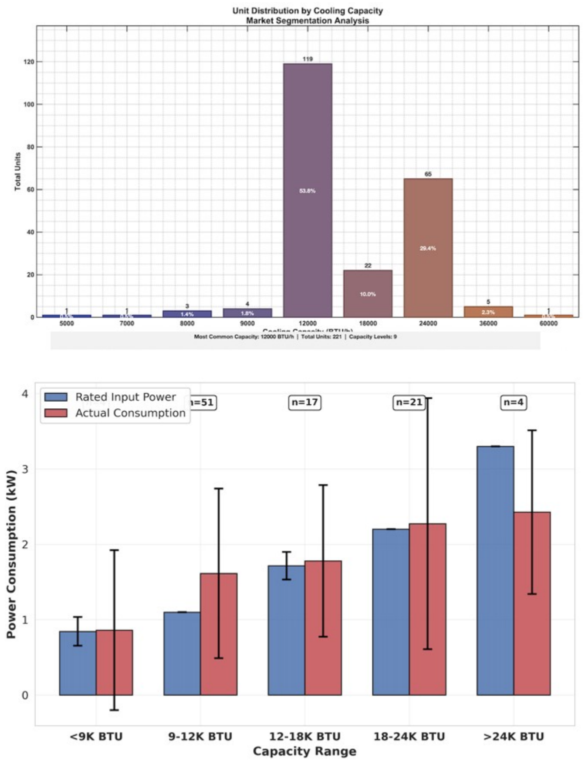

Figure 5. Distribution of BTU/h capacity in the population.

It can be observed in Fig.5 that 53.8% are 12,000 BTU/h, 29.4% are 24,000 BTU/h, and 10% are 18,000 BTU/h, totalling 93% in units between 12,000 and 24,000 BTU/h. The difference between actual consumption measurements and manufacturer datasheet specifications can also be observed.

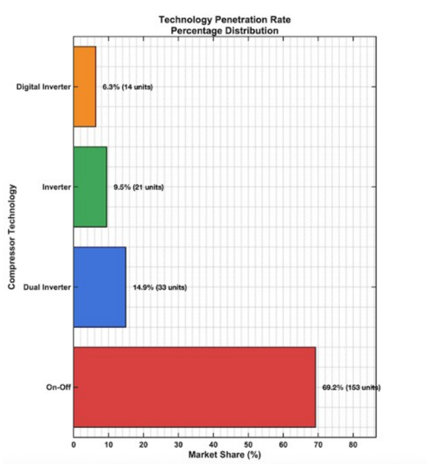

Figure 6. Percentage of technology penetration in the population of air conditioning equipment.

It can be observed in Fig.6 that 69.2% is On-Off technology, 14.9% is Dual Inverter, 9.5% is Inverter, and 6.3% is Digital Inverter.

TABLE 5
HVAC EQUIPMENT BRANDS AND THEIR AVERAGE CONSUMPTION

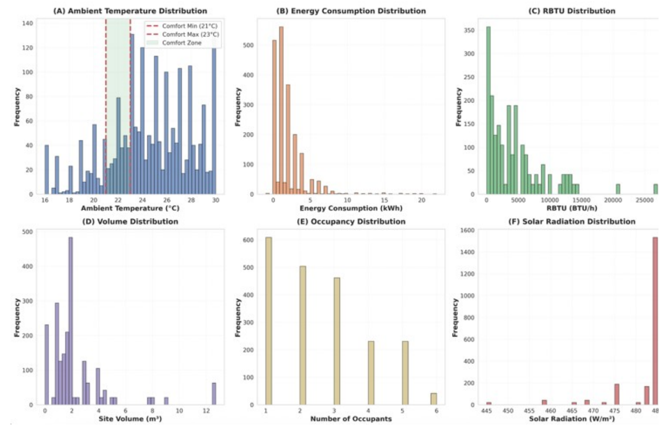

Figure 7. Statistical distributions of six key variables in the HVAC energy consumption dataset.

In Fig.7, panel (A) displays the distribution of ambient temperature (TA), with the thermal comfort zone (21-23°C) highlighted in green and bounded by red dashed lines. The distribution shows most measurements occur outside the comfort zone, with a concentration between 24-28°C. Panel (B) illustrates the energy consumption (kWh) distribution, revealing a right-skewed pattern with most values concentrated near zero and a long tail extending to higher consumption values. Panel (C) shows the RBTU (difference between installed and calculated HVAC capacity) distribution, demonstrating positive values with occasional peaks. Panel (D) depicts site volume distribution, showing a concentration of measurements for smaller spaces (40-80 m³) with fewer observations in larger volumes. Panel (E) presents occupancy distribution, indicating that most sites accommodate 2-4 occupants, with a clear discrete distribution pattern. Panel (F) displays solar radiation distribution, showing values clustered around 400-500 W/m² with a symmetric distribution.

The distributions reveal important characteristics of the dataset that influence model development. The ambient temperature distribution's skewness toward higher values suggests that HVAC systems in this study operate in cooling mode rather than heating. The concentration of measurements outside the comfort zone (21-23°C) indicates that energy consumption occurs primarily when maintaining temperature setpoints different from optimal comfort conditions. The right-skewed kWh distribution is typical of energy consumption data, where baseline consumption is common but occasional high-demand periods create outliers. The discrete nature of occupancy distribution aligns with residential and small commercial applications. The narrow range of solar radiation values (350-600 W/m²) suggests data collection occurred under similar meteorological conditions, which may limit generalizability but ensures consistency in environmental forcing variables.

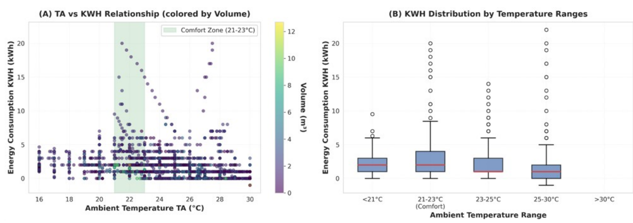

Figure 8. Analysis of the relationship between ambient temperature (TA) and energy consumption (kWh) through two complementary visualizations.

Panel (A) presents in Fig.8, a scatter plot of TA versus kWh, with data points coloured by site volume (viridis colormap ranging from purple for smaller volumes to yellow for larger volumes). The thermal comfort zone (21-23°C) is highlighted with a green vertical band. The scatter plot reveals substantial variability in energy consumption across all temperature ranges, with a general trend of higher consumption at temperatures further from the comfort zone. The colour gradient shows that larger volumes (yellow points) tend to consume more energy, though this relationship exhibits considerable scatter. Panel (B) displays box plots of kWh distribution across five temperature ranges: <21°C, 21-23°C (Comfort), 23-25°C, 25-30°C, and >30°C. The comfort zone shows the lowest median energy consumption with narrow interquartile range, while temperatures >30°C exhibit the highest median consumption and greatest variability.

The model employs a deep neural network architecture with four hidden layers of 256, 128, 64, and 32 neurons respectively, using ReLU (Rectified Linear Unit) activation functions. Advanced techniques were also implemented including L2 Regularization ($\lambda = 0.001$), Batch Normalization in all layers, Dropout (0.2-0.3) to prevent overfitting, Early Stopping (patience=50), and perturbation sensitivity analysis.

Meteorological Variables included outdoor temperature (°C), relative humidity (%), and solar radiation (W/m²). Meteorological time series were obtained from local stations with hourly resolution. Occupancy Variables included number of occupants per zone and calendar (day of week, time of day, holiday indicator). Occupancy detection was implemented using PIR (passive infrared) sensors and entry/exit counters. Operational Variables included air supply temperatures (°C), compressor types, and estimated internal thermal loads (BTU/h). Historical Features included energy consumption from the last 24 hours (kWh), 7-day moving averages, and weekly trends.

Model training used data collected from representative households, totalling 4,420 temperature and electrical consumption measurements.

E. Model Results

TABLE 6
ARTIFICIAL NEURAL NETWORK RESULTS

Dataset	RMSE	MAE	R²	Correlation
Training	1.047	0.616	0.780	0.884
Validation	0.980	0.675	0.809	0.900
Test	1.079	0.669	0.778	0.884
Comfort (21-23°C)	1.011	0.713	0.907	-

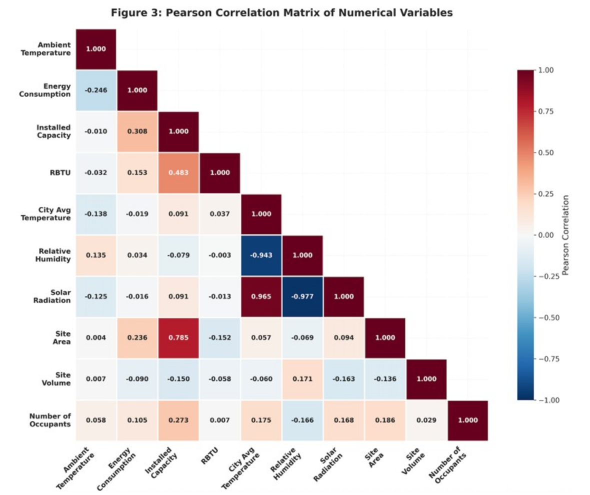

Figure 9. Triangular correlation heatmap displaying Pearson correlation coefficients between ten key numerical variables in the HVAC dataset.

The matrix uses a red-blue diverging colormap in Fig.9, where red indicates positive correlations, blue indicates negative correlations, and white represents near-zero correlation. Only the lower triangle is displayed to avoid redundancy. Correlation coefficients are annotated within each cell, providing quantitative values ranging from -1 to +1. The variables include Ambient Temperature (TA), Energy Consumption (kWh), Installed Capacity (BTUI_final), RBTU, City Average Temperature (AVT), Relative Humidity (HRC), Solar Radiation (SR), Site Area (AR), Site Volume (V), and Number of Occupants (NP). Strong positive correlations ($r > 0.5$) are evident between physically related variables such as site area and volume ($r \approx 0.8$-0.9), while most environmental variables show weak to moderate correlations with energy consumption.

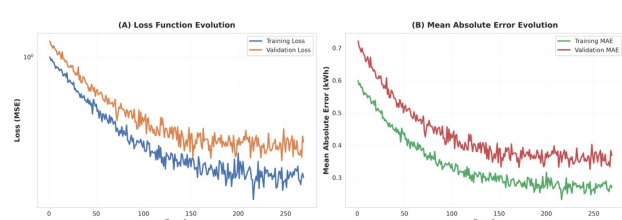

Figure 10. Dynamics of the deep neural network through two panels showing performance metrics evolution across 269 training epochs.

Panel (A) displays in Fig. 10, the loss function (Mean Squared Error, MSE) on a logarithmic scale, showing both training loss (blue line) and validation loss (orange line). Both curves exhibit typical learning behaviour: rapid initial decrease during the first 50 epochs, followed by gradual improvement and eventual plateau. The validation loss closely tracks the training loss with slightly higher values, indicating good generalization without significant overfitting. Panel (B) shows the Mean Absolute Error (MAE) evolution for both training (green line) and validation (red line) sets. The MAE curves mirror the loss patterns but on a linear scale, providing an interpretable metric in kWh units. Both training and validation MAE converge to 0.6-0.7 kWh, demonstrating consistent model performance across datasets.

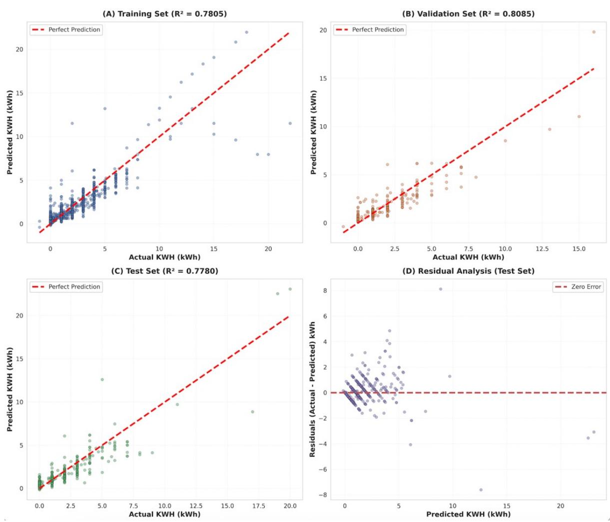

Figure 11. Comprehensive evaluation of model prediction accuracy across different dataset partitions.

Panels (A), (B), and (C) in Fig.11, present scatter plots of predicted versus actual kWh values for training, validation, and test sets, respectively. Each panel includes a red dashed line representing perfect prediction (y = x), allowing visual assessment of prediction accuracy. The scatter points cluster around this diagonal line, with some dispersion indicating prediction errors. Panel (A) shows training set performance (R^2 = 0.7805), Panel (B) displays validation set performance (R^2 = 0.8085), and Panel (C) illustrates test set performance (R^2 = 0.7780). Panel (D) presents a residual plot for the test set, showing prediction errors (actual minus predicted) plotted against predicted values. The residuals scatter around zero (indicated by a red horizontal dashed line) with symmetric distribution above and below, suggesting unbiased predictions.

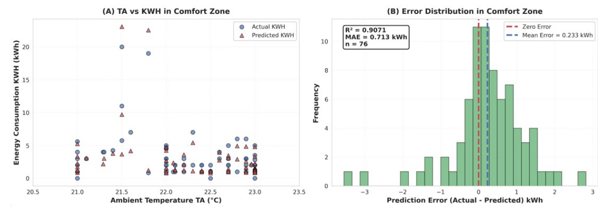

Figure 12. Model performance within the thermal comfort zone (21-23°C).

Fig. 12 focuses on which is a primary objective of this study. Panel (A) presents a scatter plot comparing actual kWh (blue circles) and predicted kWh (red triangles) against ambient temperature within the 21-23°C range. The plot is constrained to the x-axis range of 20.5-23.5°C to emphasize the comfort zone. Both actual and predicted values show considerable scatter, but predictions generally track the actual values across the temperature range. Panel (B) displays a histogram of prediction errors (actual minus predicted) within the comfort zone, revealing a normal distribution centred near zero. The distribution is overlaid with two vertical dashed lines: red at zero error (perfect prediction) and blue at the mean error (-0.232 kWh), which is remarkably close to zero. A text box in the upper left corner provides key performance metrics: R^2 = 0.9071, MAE = 0.713 kWh, and sample size n = 76.

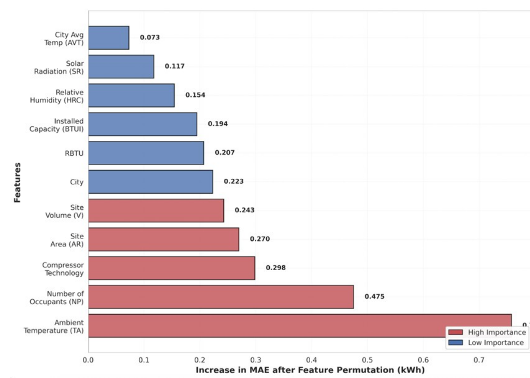

Figure 13. Quantification of the relative importance of each input feature using the permutation importance method.

The horizontal bar chart in Fig. 13 displays eleven features ranked by their impact on model performance, measured as the increase in Mean Absolute Error (MAE) when each feature is randomly permuted (shuffled). Features are color-coded: red bars indicate high importance (above median), while blue bars indicate lower importance. Ambient Temperature (TA) dominates with an important score of approximately +0.958 kWh increase in MAE, nearly double the second-most important feature, Number of Occupants (NP, +0.458 kWh). The remaining features show progressively smaller importance values, with City Average Temperature (AVT) contributing least (+0.033 kWh). Each bar is annotated with its precise numerical value for reference.

The results demonstrate firstly that nameplate consumption data are only reference data, and that HVAC equipment of the same BTU/h capacity, model, and brand can have different energy consumption. Table 7 shows the average consumption of all equipment cited in this study for reaching the comfort temperature zone.

TABLE 7
ENERGY CONSUMPTION AT COMFORT ZONE TEMPERATURES

Temperature	Average Consumption
21°C	2.37 ± 1.37 kWh
22°C	3.38 ± 4.81 kWh
23°C	0.87 ± 1.23 kWh

IV. CONCLUSIONS

This article provides a strategy for climate control based on a model generated by an artificial neural network whose output variable has been designed to serve as an input variable in an intelligent control system for HVAC systems. This modelling is developed from real data, and it was observed that manufacturer consumption data differs from actual consumption. This is because manufacturer data are nominal values under conditions different from actual conditions, as there are many variables that can affect HVAC equipment performance.

This study finds that the most efficient brands are Electrolux (1.17 kWh), Panasonic (1.35 kWh), and TCL (1.44 kWh). The least efficient are ECASA (2.95 kWh) and SMC (2.54 kWh, with high variability). Capacity segmentation for the residential sector shows <18K BTU/h (majority of the dataset). Average actual consumption is 1.871 kW, average specification is 1.567 kW, average ratio is 1.261 (26.1% overload), and median is 1.082 (8.2% overload). This study also finds that forty-two equipment units are more efficient than manufacturer specifications (42.9%), while fifty-six equipment units are less efficient than manufacturer specifications (57.1%).

The proposed neural network constitutes a robust and reproducible mathematical model for energy consumption prediction, integrating variables from equipment technical characteristics such as BTU/h capacity, meteorological variables such as temperature, humidity, and solar radiation in the city, and thermal load variables such as area, number of people, and other thermal loads. Its structure allows future extension toward intelligent predictive HVAC control and IoT systems [7]. The trained neural network achieved the following metrics: R^2 of 0.778 (the model explains 77.8% of variability in kWh), MAE of 0.669 kWh (average error of only 0.67 kWh), and RMSE of 1.079 kWh (root mean square error). Performance in the Comfort Zone of 21 to 23°C achieved 90.7% accuracy. Average predicted consumption is 2.77 kWh across all equipment types, compressor technologies, brands, models, and capacities.

TA is the most important predictor (+0.958 kWh impact). Actual consumption is 26% higher than factory specifications. Only 18 of 98 equipment units (18.4%) meet specifications (±10%). These results confirm the viability of using artificial neural networks as predictive engines for HVAC equipment energy consumption, and the objective of calculating the kWh variable to reach temperatures in the comfort zone is achieved.

This output can be used as input to an MPC controller that also has the thermal comfort variable as input to optimize and find a balance between comfort and energy consumption. The framework is scalable, suitable for real-time IoT implementation, and applicable to residential and commercial buildings in tropical regions to achieve energy savings, comfort improvement, and scalability.

REFERENCES

[1] Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2014). A comprehensive overview on the data driven and large scale-based approaches for forecasting of building energy demand: A review. Energy and Buildings, 165, 301-320.

[2] Drgona, J., Arroyo, J., Cupeiro Figueroa, I., Blum, D., Arendt, K., Kim, D., ... & Helsen, L. (2018). All you need to know about model predictive control for buildings. Annual Reviews in Control, 50, 190-232.

[3] Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., ... & Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. Energy and Buildings, 45, 15-27.

[4] Ma, Y., Borrelli, F., Hencey, B., Coffey, B., Bengea, S., & Haves, P. (2012). Model predictive control for the operation of building cooling systems. IEEE Transactions on Control Systems Technology, 20(3), 796-803.

[5] ASHRAE. (2020). ASHRAE Standard 55: Thermal environmental conditions for human occupancy. American Society of Heating, Refrigerating and Air-Conditioning Engineers.

[6] International Organization for Standardization. (2004). ISO 8996: Ergonomics of the thermal environment - Determination of metabolic rate. ISO.

[7] Bolanos, J. J. L., Guingla, D. A. P., Quintero, C. A. C., & Casal, J. J. L. (2025). IoT-based HVAC system modelling. In E3S Web of Conferences (Vol. 658, p. 03004). EDP Sciences.

[8] Rahman, A., Srikumar, V., & Smith, A. D. (2017). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, 212, 372-385.

[9] Sturzenegger, D., Gyalistras, D., Morari, M., & Smith, R. S. (2016). Model predictive climate control of a Swiss office building: Implementation, results, and cost-benefit analysis. IEEE Transactions on Control Systems Technology, 24(1), 1-12.

[10] Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. Sustainable Cities and Society, 39, 508-518.

[11] Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394-398.

[12] Afram, A., & Janabi-Sharifi, F. (2014). Theory and applications of HVAC control systems – A review of model predictive control (MPC). Building and Environment, 72, 343-355.

[13] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192-1205.

[14] Singh, J., Gupta, M., & Kumar, R. (2023). Smart technologies and intelligent systems for HVAC applications. Energy Reports, 9, 4145-4162.

[15] Mahboobtosi, M., Ehsani, H., Ganji, A. M., Jalili, P., & Hassan, M. H. (2024). Performance analysis of variable speed compressor technologies in residential air conditioning systems. Applied Thermal Engineering, 236, 121549.

[16] Jeon, Y. J., Park, S. J., Lee, H., Cui, J. S., & Jung, D. H. (2023). Emulating expert evaluation methods for building energy performance using neural network approaches. Energy and Buildings, 278, 112612.

[17] International Energy Agency. (2019). Global Status Report for Buildings and Construction 2019. IEA Publications.

[18] Phetsuwan, K., & Pora, W. (2022). A direct load control algorithm for air conditioning systems based on coefficient of performance optimization. Energy and Buildings, 271, 112345.

[19] Sambasivarao, K. V., & Bhima, A. S. R. D. (2023). Artificial intelligence and computational methods for HVAC system optimization: A comprehensive review. Journal of Building Engineering, 65, 105742.

[20] Instituto Nacional de Meteorología e Hidrología (INAMHI). (2023). Anuario Meteorológico del Ecuador 2023. Quito, Ecuador: INAMHI.