Advances in Energy Science and Technology

Vol. 1, No. 1|1 June 2025

AEST

Application of random forest algorithm in identification of key factors in building cooling system energy consumption and energy conservation strategies (for office-educational buildings in Amirkabir University of Technology)

Romina Poursedigh[a], Farzad Hatami[b,*], Eghbal Shakeri[a]

[a] *Civil and Environmental Engineering Faculty, Amirkabir University of Technology*

[b] *Structural and Earthquake Research Institute, Amirkabir University of Technology*

Keywords

Building Energy Consumption
Optimization
Random Forest Algorithm
Occupants' Behavior
Sustainable development

Article Info

DOI: 10.22060/AEST.2025.5751

Received 6 Jan 2025

Accepted 1 Feb 2025

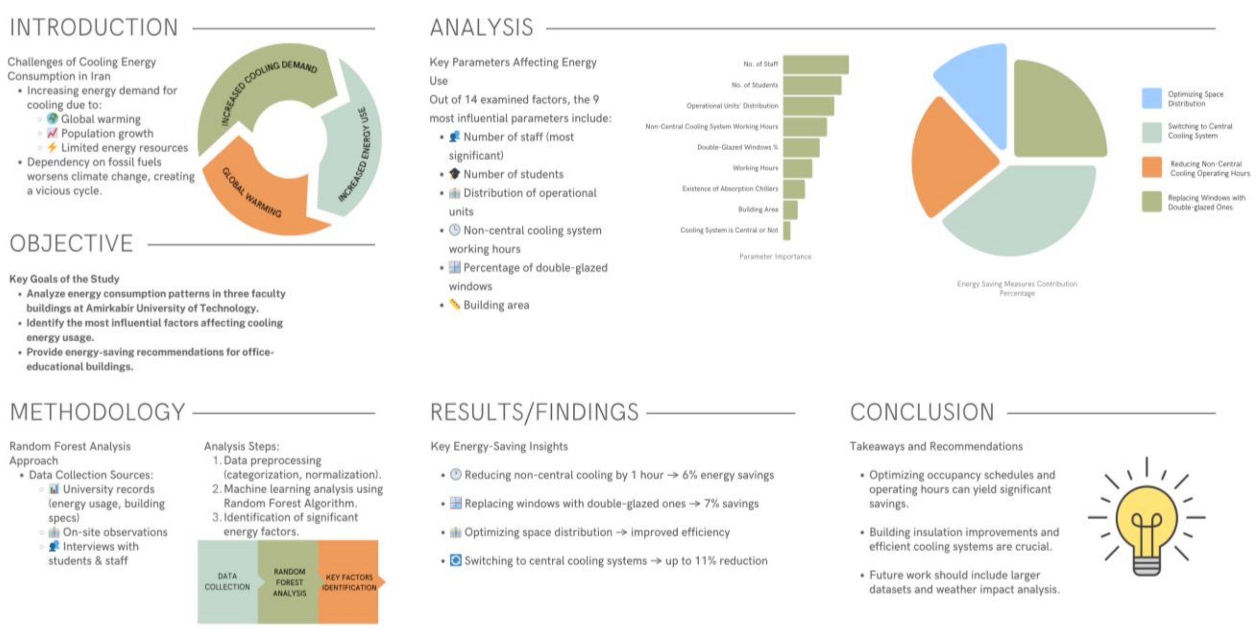
Published 1 June 2025

* **Corresponding author:**
Hatami@aut.ac.ir

Abstract

The increasing electricity consumption for air cooling and conditioning in buildings has become a significant challenge in Iran, driven by population growth, global warming, limited energy production and distribution capacities, and concerns about the reliance on fossil fuels for electricity generation. This study investigates energy consumption patterns of cooling systems in three buildings at Amirkabir University of Technology: Civil and Environmental Engineering (Building No. 2), Computer Engineering, and Aerospace Engineering. Using energy consumption data, the study identifies key factors related to building characteristics and user behavior that impact cooling energy usage. A Random Forest analysis revealed that among 14 factors, 9 were most significant, with the number of staff, number of students, and the distribution percentage of operating units ranked as the top three. These findings provide insights for developing effective energy conservation strategies tailored to university buildings in similar contexts.

Graphical Abstract



1. Introduction

Energy consumption represents one of the most critical challenges to achieving sustainable development in today's world [1,2]. Population growth, urbanization, global warming, and the depletion of fossil fuel resources pose significant obstacles to providing electricity for building cooling systems [3,4]. Moreover, the increase in energy consumption, coupled with the burning of fossil fuels to meet this demand, exacerbates climate change, leading to natural disasters such as tsunamis, floods, and wildfires[3-5]. This, in turn, increases the demand for cooling, creating a destructive feedback loop that undermines environmental sustainability and development goals [6].

In addition to transitioning from non-renewable to renewable energy sources, reducing energy consumption—particularly in buildings—is essential[7]. Understanding the factors influencing energy use in buildings can provide policymakers and decision-makers with valuable insights for designing effective energy management strategies [8]. This approach helps to avoid the inefficiencies and costs associated with trial-and-error implementation of energy-saving measures[9,10].

This study focuses on identifying the most significant factors influencing the energy consumption of cooling systems in office-educational buildings at Amirkabir University of Technology. Data were collected from the university's mechanical facilities office, as well as through field observations and interviews. A Random Forest (RF) model, implemented in Python, was used to analyze the data, and solutions were proposed to reduce energy usage.

The subsequent sections of this paper detail the research process. The "Data Collection" section describes the types and methods of gathering electricity consumption data from Amirkabir University. The "Methodology" section outlines the data preprocessing steps and the Random Forest model used for analysis. The results of the analysis, along with proposed energy-saving solutions, are presented in the fourth section. Finally, the study concludes with a discussion of its limitations and recommendations for future research.

Data collection

Building energy consumption data can be categorized into three groups: (1) physical attributes of the building and its energy-consuming systems, (2) weather data, and (3) occupants' characteristics [11].

The physical attributes of buildings and their energy-consuming systems encompass factors such as building age, area, materials, insulation properties, and the characteristics of systems like heating, cooling, conditioning, and lighting. Weather data, which include variables such as air temperature, natural lighting intensity and duration, wind speed, and other environmental factors, significantly influence the energy consumption of systems in the first category. The third group, referred to as occupancy data, represents the characteristics and behaviors of building occupants. These data are the primary cause of the "performance gap" between energy consumption estimates made by prediction models and the actual energy use of buildings [12]. Occupancy data are typically complex, difficult to obtain, and cannot be incorporated into energy consumption estimates using purely physical calculations [6,9].

Collecting comprehensive and accurate data for all three categories is a challenging and time-consuming task [11,13]. While reducing the number of parameters used in energy consumption estimation may decrease prediction accuracy, it can significantly reduce the costs associated with data collection, prediction calculations, and energy management. Furthermore, not all data parameters contribute equally to energy usage. Identifying the most critical parameters can improve calculation efficiency and help determine the most effective energy conservation strategies [14,15].

1.1. Data collection from Amirkabir University of Technology

Energy consumption data from the Civil and Environmental Engineering (Building No. 2), Computer Engineering, and Aerospace Engineering faculties were collected from the university's mechanical facilities office. These data include energy consumption, building area, cooling system type (central or non-central), cooling system sub-type, and cooling system age. The data span three consecutive summer days in 2022, from September 22 to September 24. Additional parameters, identified through specialist opinions and a literature review, include the distribution of operational units, percentage of double-glazed windows, staff work hours, professors' and students' presence duration, and the daily number of individuals present in each building.

The number of students present in each faculty building was determined through interviews with students, computer room operators, and library and study room staff. Since no classes are held during summer, the student count reflects those utilizing laboratories, study rooms, and computer rooms, representing the average number of individuals present per day.

The number of staff present daily in each building was collected, with privacy protections in place, from the staff management office of each faculty. Work hours represent the average duration of staff presence in the buildings. The number of professors present each day was assumed to be half of the total number of professors assigned to each faculty.

The percentage of double-glazed windows reflects the ratio of double-glazed windows to the total number of windows in each faculty, as reported by the university's mechanical facilities office.

The distribution percentage of operational units was gathered following a suggestion from the university's mechanical facilities office. This parameter represents the ratio of floors with rooms used for offices, laboratories, professors' offices, computer rooms, and study halls to the total number of floors requiring cooling during summer (excluding classrooms, which were assumed to not require cooling during the observation period). This parameter is particularly significant when studying non-central cooling systems and was collected through field observations and interviews with staff and students.

1.2. Approximation of daily average duration of presence of professors and students

Since the presence of professors and students is not recorded during the summer, their average duration of presence was assumed to be equal to the staff's working hours. However, this parameter is inherently subject to some degree of error.

1.3. Approximation of the age of buildings and cooling systems

The age of each building is calculated as an approximate average of the ages of its floors, taking into account renovations and the construction of additional floors over the building's lifespan. Similarly, the age of the cooling systems is determined as an average, factoring in total overhauls and repairs. Both parameters were obtained through inquiries made to the university's mechanical facilities office.

1.4. Approximation of working hours of central and non-central cooling systems

To ensure thermal comfort for occupants and considering the operational characteristics of central cooling systems, the "pre-cooling" process typically begins approximately one hour before people arrive at the building. It is assumed that these systems are turned off immediately after occupants leave the building. Consequently, the operating duration of central cooling systems is considered to be the daily working hours plus one additional hour. In contrast, non-central cooling units are presumed to be directly controlled by occupants and to operate only while people are present in the room, corresponding to the staff's working hours.

1.5. Translating Computers' Cooling Requirements into Human Equivalents

Assuming that laboratory machines and equipment are adequately insulated, the heat generated by computers (both university-owned and students' personal devices) contributes to the cooling load on the building's cooling system. According to the literature [16,17], each laptop generates a cooling load equivalent to that of an average human, while each desktop computer produces a cooling load equivalent to that of three average humans. Consequently, a person with a laptop imposes a cooling load equivalent to two people, and a person using a desktop computer imposes a cooling load equivalent to four people. Additionally, some desktop computers were operated remotely by students, generating heat equivalent to three people present in the building. The number of desktop computers used remotely or within university buildings, as well as the number of individuals present with laptops, was determined through field observations and interviews.

2. Methodology

As previously mentioned, incorporating occupant behavior into energy consumption prediction solely through physical calculations is time-consuming or even impractical. Therefore, this study employs a Random Forest model to:

- Establish a method for estimating energy consumption in office-educational buildings, and
- Identify the most influential parameters affecting cooling system energy usage based on the collected data.

This section describes the Decision Tree and Random Forest algorithms, along with preprocessing steps, which are essential before introducing data to a machine learning regression model.

2.1. Decision Tree

A Decision Tree is a supervised machine learning algorithm that uses data entropy to categorize information based on conditions that maximize "Data Gain." "Figure 1" illustrates a schematic representation of a Decision Tree with two independent parameters.

As shown in "Fig. 1", the Decision Tree is trained using a dataset with two independent parameters (X_1 and X_2). The tree begins categorizing data by imposing conditions on these parameters. For instance, in the first step, drawing the line X2=a.results in data on each side of the line being more similar compared to any other hypothetical line. The tree then draws subsequent lines (X1=b and X1=c) on each side of X2=a to achieve the same goal iteratively. This process continues until the regression error falls below a target threshold or a predefined number of iterations is reached. Finally, the model assigns the input data sample to one of the established categories based on the training conditions, and the predicted output is the average of outputs in that category.

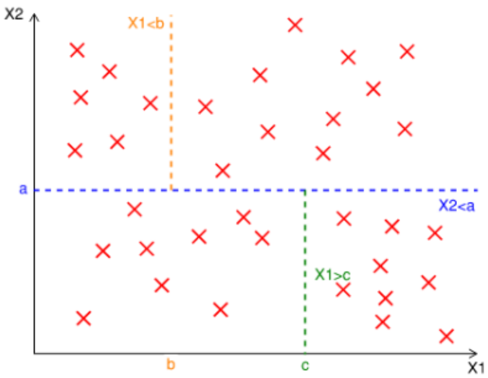


Fig. 1. An example of a decision tree.

2.2. Random Forest

A Random Forest is an ensemble learning algorithm comprising multiple Decision Trees [18]. The Random Forest output is the average of outputs from its individual Decision Trees, each incorporating different subsets of input parameters. This structure makes Random Forest more accurate than a single Decision Tree, less prone to overfitting, and more effective at ranking input parameter significance. All parameters have multiple opportunities to participate in calculations, enhancing the model's robustness [19, 20].

2.3. Preprocessing

Preprocessing is a critical step in developing machine learning algorithms. It typically involves the following tasks:

- Separating input parameters from the output(s),
- Quantifying categorical data,
- Normalizing parameter scales,
- Dividing datasets into “Learning” and “Testing” datasets.

2.3.1. Separating Input Parameters from the Output(s)

In supervised machine learning, independent (input) and dependent (output) parameters are introduced to the model. The algorithm learns the relationship between them and uses this understanding to predict outputs for new input data. To achieve this, it is essential to clearly identify the “goal” parameter to guide the model.

2.3.2. Quantifying Categorical Data

Qualitative data cannot be directly compared to quantitative data in a model. Therefore, it must be transformed into numerical form. For binary categories (e.g., Yes/No), values such as 0 and 1 can be assigned. Similarly, categorical data with “n” categories can be represented by integers from 0 to “n−1”. However, including all “n” categories introduces redundancy—known as the “dummy variable trap.” For instance, a parameter with three categories (0, 1, or 2) does not require explicitly representing “n−1”, as the third category can be inferred.

2.3.3. Normalizing Parameter Scales

Parameters with different ranges or amplitudes must be scaled to ensure equal influence on output predictions. For instance, a parameter ranging from 0 to 1 could otherwise be overshadowed by one ranging from 10 to 1000. Standardization is commonly used, transforming parameters to have an average of 0 and a variance of 1. Importantly, the specific average and variance values do not affect the prediction as long as they are consistent across all parameters.

2.3.4. Separating Datasets into Learning and Testing Sets

Machine learning models are trained on “Learning” datasets and evaluated on “Testing” datasets to measure performance. A typical ratio for splitting data is 70:30 between Learning and Testing sets.

After preprocessing, the data is prepared for introduction into the machine learning algorithm.

3. Results

The primary objective of this study was to identify the most influential factors affecting cooling system energy consumption and to determine the most effective energy conservation strategies for office-educational buildings at Amirkabir University of Technology.

Using a Random Forest model, the collected data were analyzed, revealing that 9 out of 14 parameters significantly influenced building cooling system energy usage. Fig. 2 illustrates the changes in mean squared error (MSE) after systematically removing each parameter. The “jumps” in the diagram indicate that the removal of 5 specific parameters does not significantly impact the model’s accuracy. These parameters are:

- Cooling system age,
- Central cooling system working hours,
- Existence of evaporative coolers,
- Building age, and
- Number of professors.

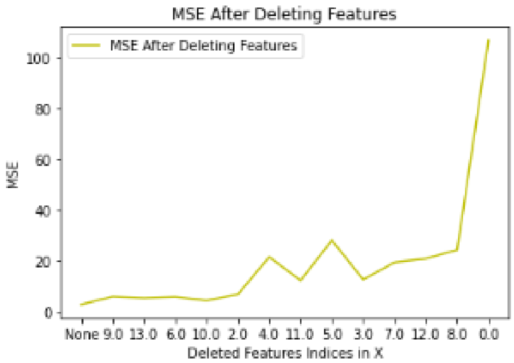


Fig. 2. MSE Changes after parameter omissions.

The remaining 9 parameters, identified by the Random Forest model as the most influential in cooling system energy consumption, are ranked below in order of importance, from most to least significant:

- Number of staff,
- Number of students,
- Distribution of operational units,
- Non-central cooling system working hours,
- Percentage of double-glazed windows,
- Average working hours,
- Existence of absorption chillers,
- Building area, and
- Whether or not the cooling system is central.

3.1. Effect of Staff Presence on Cooling System Energy Usage

Table 1. Hourly Staff Numbers in Civil and Environmental Engineering Faculty.

Time \ Date	Sep 22	Sep23	Sep24
6-7	3	3	5
7-8	13	13	10
8-9	13	15	14
9-10	14	15	14
10-11	14	15	14
11-12	14	15	14
12-13	14	15	13

13-14	14	15	13
14-15	9	8	6
15-16	7	7	4
16-17	7	6	4
17-18	6	4	4
18-19	5	4	2
19-20	4	1	0
20-21	3	0	0
21-22	1	0	0

The number of staff present each day emerged as the most influential factor in building cooling system energy consumption. Staff presence records (provided without personal identifiers) were used to calculate the number of staff present each hour, as shown in "Tables 1–3". Regression results, considering only staff numbers, are illustrated in "Figs. 3–5".

It was observed that most staff members left their respective faculties by 3–4 p.m. Additionally, the number of students and professors was assumed constant throughout the day, leading to inaccuracies in regression results after 4 p.m. Despite this, other parts of the diagrams confirm a direct correlation between energy usage and staff numbers, aligning with findings in existing literature [21].

Table 2. Hourly Staff Numbers in Aerospace Engineering Faculty.

Date / Time	Sep 22	Sep23	Sep24
6-7	1	3	0
7-8	7	7	4
8-9	8	7	4
9-10	8	7	4
10-11	8	7	4
11-12	8	7	4
12-13	8	7	4
13-14	8	7	4
14-15	6	5	4
15-16	5	5	3
16-17	5	5	3
17-18	2	1	1
18-19	1	1	1
19-20	1	1	0
20-21	1	0	0
21-22	1	0	0

Advances in Energy Science and Technology
Vol. 1, No. 1 | 1 June 2025
AEST

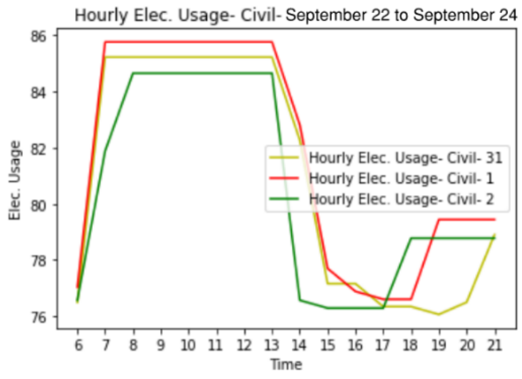


Fig. 3. Hourly Electricity Usage in Civil and Environmental Engineering Faculty.

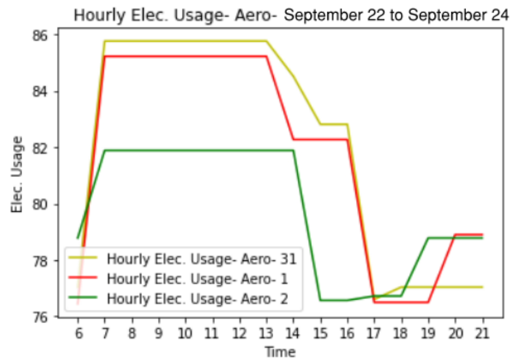


Fig. 4. Hourly Electricity Usage in Aerospace Engineering Faculty.

3.2. Effect of Distribution of Operational Units on Cooling System Energy Usage

Table 3. Hourly Staff Numbers in Computer Engineering Faculty.

Time \ Date	Sep 22	Sep23	Sep24
6-7	3	1	1
7-8	7	4	4
8-9	7	6	5
9-10	7	6	5
10-11	7	6	5
11-12	7	6	5
12-13	7	6	5
13-14	7	6	5
14-15	6	4	3
15-16	4	4	3
16-17	4	3	3
17-18	2	3	2
18-19	2	2	2
19-20	2	2	1
20-21	1	0	0
21-22	0	0	0

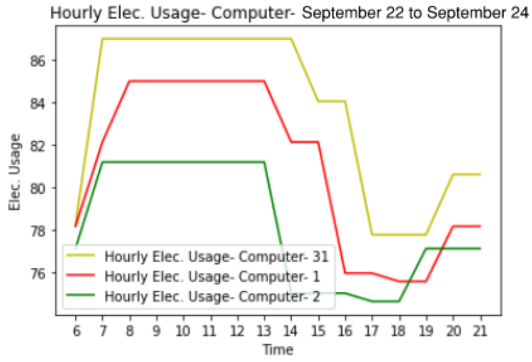


Fig. 5. Hourly Electricity Usage in Computer Engineering Faculty.

During summer, with no classes being held, it was assumed that professors' offices, laboratories, study halls, and computer rooms were the primary spaces requiring cooling. Consequently, the distribution of these spaces across building floors significantly affected cooling system energy consumption.

At the time of data collection, the following floors required cooling:

- Civil and Environmental Engineering faculty: 4 out of 7 floors,
- Aerospace Engineering faculty: 4 out of 5 floors,
- Computer Engineering faculty: 3 out of 4 floors.

Assuming an optimal distribution (e.g., two floors per building: one for professors' offices and one for other operational units), the estimated reduction in energy consumption was less than 0.3% per building. Although ideal, this scenario is not necessarily practical.

3.3. Effect of Reducing Non-Central Cooling System Operating Hours

According to the Random Forest model, reducing the operating duration of non-central cooling systems by 60 minutes could lead to a reduction of up to 6% in energy consumption. The importance of this parameter has also been highlighted in related studies [22].

3.4. Increasing the Percentage of Double-Glazed Windows

Replacing all windows in the studied buildings with double-glazed ones could result in a reduction of up to 7% in energy consumption. Additionally, the window-to-wall area ratio on perimeter walls significantly impacts energy usage. As noted in [23], this ratio's optimal value depends on wall orientation, but an average ratio of 0.2 could potentially reduce annual cooling loads by 15–20 kWh per square meter of building area.

3.5. Effect of Absorption Chillers and System Centralization

To evaluate the impact of absorption chillers on energy consumption, the cooling systems were hypothetically altered:

- Civil and Environmental Engineering faculty (originally relying on absorption chillers) was modeled with non-central cooling systems.
- Aerospace and Computer Engineering faculties were modeled with combined (central and non-central) cooling systems.

The results revealed the following changes:

- Civil and Environmental Engineering faculty: up to 11% reduction in energy consumption.
- Aerospace Engineering faculty: up to 17% increase in energy consumption.
- Computer Engineering faculty: up to 18% increase in energy consumption.

These findings align with prior studies, which indicate that giving occupants control over HVAC systems can lead to up to 70% energy savings, while improper usage may increase consumption by up to 35% [24,25].

Limitations:

Data Sample Size:

- The study analyzed data from only a few days, limiting accuracy. With more samples, results could vary, and the significance of certain parameters, such as building area, cooling system age, and building age, may become more evident.
- Literature suggests building area as a key parameter in energy usage [21,22].

System and Building Age:

- Cooling systems across all buildings were older than their optimal lifecycle, affecting efficiency. Updated systems might yield different results, with newer central systems potentially outperforming non-central ones [26].
- Building age, identified as a critical factor in energy usage in other studies [21], may have been underestimated due to the similar ages of the buildings analyzed.

Building Insulation:

- None of the buildings had proper insulation. Literature suggests adequate insulation could reduce energy consumption by up to 64%, especially when paired with passive cooling measures [21,26,27].

Occupancy Details:

- Parameters such as professors’ presence may have been underestimated due to a lack of detailed data. Professors directly control their cooling systems, which significantly impacts energy consumption.

Occupants' Characteristics:

- Privacy concerns prevented collecting personal data such as energy habits, thermal preferences, or demographic information.

Weather Factors:

- Data collection was limited to three days with similar weather conditions, excluding significant weather variability's impact on cooling energy usage.

Other Building Systems:

- The study did not consider heat production and energy consumption from lighting or laboratory equipment due to measurement constraints.

4. Conclusions

This study identified the most influential factors in cooling system energy usage in office-educational buildings at Amirkabir University of Technology. While limitations affected some parameters' evaluation, the results offer valuable insights into energy conservation strategies, emphasizing the importance of staff presence, cooling system operation hours, and building insulation. Future studies could benefit from larger datasets, a broader scope of parameters, and improved data collection methods to enhance analysis accuracy.

Acknowledgments

Completion of this research wouldn’t have been possible without the assistance and guidance of Mechanical Facilities and also Staff management offices of the university in the process of data collection.

The authors would further like to thank students and staff of Civil and Environmental Engineering, Computer Engineering and Aerospace Engineering faculties of Amirkabir University of Technology for their direct or indirect contributions to this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] 'Energy union'. https://energy.ec.europa.eu/topics/energy-strategy/energy-union_en (accessed Apr. 17, 2023).

[2] 'Environmental and Energy Study Institute | Ideas. Insights. Sustainable Solutions.' https://www.eesi.org/ (accessed Apr. 17, 2023).

[3] A. Habibollahzade, E. Gholamian, P. Ahmadi, A. Behzadi, Multi-criteria optimization of an integrated energy system with thermoelectric generator, parabolic trough solar collector and electrolysis for hydrogen production, Int. J. Hydrogen Energy 43 (2018) 14140–14157.

[4] U.N. Environment, 2019 Global Status Report for Buildings and Construction Sector, UNEP - UN Environment Programme, 2019. http://www.unep.org/resources/publication/2019-global-status-report-buildings-and-construction-sector (accessed Apr. 17, 2023).

[5] 'World Shale Gas Resources: An Initial Assessment of 14 Regions Outside the United States'. https://www.eia.gov/analysis/pdfpages/worldshalegasindex.php (accessed Apr. 17, 2023).

[6] A.P. Melo, R.S. Versage, G. Sawaya, R. Lamberts, A novel surrogate model to support building energy labelling system: A new approach to assess cooling energy demand in commercial buildings, Energy Build. 131 (2016) 233–247.

[7] M.M. Tahmasebi, S. Banihashemi, M.S. Hassanabadi, Assessment of the variation impacts of window on energy consumption and carbon footprint, Procedia Eng. 21 (2011) 820–828.

[8] C. Koo, S. Park, T. Hong, H.S. Park, An estimation model for the heating and cooling demand of a residential building with a different envelope design using the finite element method, Appl. Energy 115 (2014) 205–215.

[9] U. Ali, M.H. Shamsi, F. Alshehri, E. Mangina, J. O'Donnell, Application of intelligent algorithms for residential building energy performance rating prediction, in: Corrado, V., Fabrizio, E., Gasparella, A., Patuzzi, F. (Eds.), Build. Simul. 2019, 2019.

[10] P. van den Brom, A. Meijer, H. Visscher, Performance Gaps in Energy Consumption: Household Groups and Building Characteristics, *A+BE Archit.* Built Environ. 3 (2020) 03.

[11] H. Do, K.S. Cetin, Residential Building Energy Consumption: A Review of Energy Data Availability, Characteristics, and Energy Performance Prediction Methods, Curr. Sustain. Renew. Energy Rep. 5 (2018) 76–85.

[12] M.S. Kaddory Al-Zubaidy, A literature evaluation of the energy efficiency of leadership in energy and environmental design (LEED)-certified buildings, Am. J. Civ. Eng. Archit. 3 (2015) 1–7.

[13] Abrams Environmental Law Clinic, Freeing energy data—a guide for regulators to reduce one barrier to residential energy efficiency, Univ. Chicago Law School (2016).

[14] R. Langheim, G. Arreola, C. Reese, Energy efficiency motivations and actions of California solar homeowners, Proc. 2014 ACEEE Summer Study Energy Eff. Build. 7 (2014) 147–159.

[15] V. Udalov, J. Perret, V. Vasseur, Environmental motivations behind individuals' energy efficiency investments and daily energy-saving behaviour: Evidence from Germany, the Netherlands, and Belgium, Int. Econ. Econ. Policy 14 (2017) 481–499.

[16] RM Support, Technical article, *RM Support* (n.d.). Available at: https://support.rm.com/technicalarticle.asp?cref=tec455664.

[17] R. Riemer, A. Shapiro, Biomechanical energy harvesting from human motion: Theory, state of the art, design guidelines, and future directions, J. Neuroeng. Rehabil. 8 (2011) 1–13.

[18] M.B. Kursa, W.R. Rudnicki, The All Relevant Feature Selection using Random Forest, arXiv (2011).

[19] Z. Wang, Z. Zhao, B. Lin, Y. Zhu, Q. Ouyang, Residential heating energy consumption modeling through a bottom-up approach for China's Hot Summer–Cold Winter climatic region, Energy Build. 109 (2015) 65–74.

[20] R.A. Romero, G. Bojórquez, M. Corral, R. Gallegos, Energy and the occupant's thermal perception of low-income dwellings in hot-dry climate: Mexicali, México, Renew. Energy 49 (2013) 267–270.

[21] A. Alshibani, O.S. Alshamrani, ANN/BIM-based model for predicting the energy cost of residential buildings in Saudi Arabia, J. Taibah Univ. Sci. 11 (2017) 1317–1329.

[22] C. Deb, S.E. Lee, M. Santamouris, Using artificial neural networks to assess HVAC-related energy saving in retrofitted office buildings, Sol. Energy 163 (2018) 32–44.

[23] S. Zhao, Using artificial neural network and WebGL to algorithmically optimize window wall ratios of high-rise office buildings, J. Comput. Des. Eng. 8 (2021) 638–653.

[24] Z. Deng, Q. Chen, Simulating the impact of occupant behavior on energy use of HVAC systems by implementing a behavioral artificial neural network model, Energy Build. 198 (2019) 216–227.

[25] R. Gupta, M. Gregg, Empirical evaluation of the energy and environmental performance of a sustainably-designed but under-utilized institutional building in the UK, Energy Build. 128 (2016) 68–80.

[26] S.A. Sharif Arani, Optimizing energy performance of building renovation using traditional and machine learning approaches, Doctoral Dissertation, Concordia Univ. (2020).

[27] A. Al-Sakkaf, E.M. Abdelkader, S. El-Zahab, A. Bagchi, T. Zayed, A holistic methodology for lifecycle energy consumption of heritage buildings, J. Phys. Conf. Ser. 1900 (2021) 012014.

[28] A. Tech report, Tech report title, *Prog. Rep. 1, Univ. Higher Educ., Cambridge, MA* (2003). Available at: http://www.abc.edu.