12-2021

A Machine Learning-Based Approach to Predict and Optimize the Performance of Zero Energy Building (ZEB): A Case Study for Florida

Benjamin Kubwimana

A Machine Learning-Based Approach to Predict and Optimize the Performance of Zero Energy Building (ZEB): A Case Study for Florida

by

Benjamin Kubwimana

A thesis submitted to the Department of Mechanical and Civil Engineering of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Mechanical Engineering

Melbourne, Florida
December 2021

We, the undersigned committee, hereby approve the attached thesis,
“A Machine Learning-Based Approach to Predict and Optimize the Performance of Zero Energy Building (ZEB): A Case Study for Florida”
by
Benjamin Kubwimana

Hamidreza Najafi, Ph.D.
Associate Professor
Mechanical and Civil Engineering
Major Advisor

Troy Nguyen, Ph.D. P.E
Associate Professor
Mechanical and Civil Engineering

Aldo Fabregas Ariza, Ph.D.
Assistant Professor
Computer Engineering and Sciences

Ryan T White, Ph.D.
Assistant Professor
Mathematical Sciences

Ashok Pandit, Ph.D.
Professor and Department Head
Mechanical and Civil Engineering

Abstract

Title: A Machine Learning-Based Approach to Predict and Optimize the Performance of Zero Energy Building (ZEB): A Case Study for Florida

Author: Benjamin Kubwimana

Major Advisor: Hamidreza Najafi, Ph.D.

Machine learning is currently one of the most searched fields aiming to solve real-life problems. Building simulation software tools help engineers estimate building energy behaviors before the actual construction, allowing implementation of more energy efficient choices in building design and construction. Current building energy simulation software tools are mostly physics-based and still lack the benefit obtained with machine learning-based modeling, which offers fast and less computationally expensive techniques to build energy models and efficiently perform design optimization. This thesis presents a machine learning-based approach for building energy modeling and optimizing design parameters to minimize building's energy consumption. The study is comprised of three main stages. These include creating an EnergyPlus simulation model to generate a physics-based model for the building with all building characteristics. The model is used to generate a database containing input design parameters that are used in energy modeling and annual energy consumptions for different energy models. The results obtained from this database are then used in the second stage, which involves developing an artificial neural network-based surrogate model. The neural network performs simulations by taking a set of inputs and trying to predict an output. The inputs, in this case, are building design parameters and control settings, while the output is the building energy consumption, photovoltaic system power production, and the corresponding net site energy. The third stage is the optimization stage implemented on the surrogate model to determine optimal design variables that provide minimal energy consumption. Design parameter search space along with the surrogate model are provided as inputs to the optimization algorithm. The study uses two different optimization approaches, including the genetic algorithm and the Bayesian method. This

study shows that the proposed machine learning-based strategy accurately estimates overall energy usage and production. Furthermore, the model optimization is implemented on the neural network at far less computational costs and time than the traditional strategies that involve numerous co-simulation tools to obtain the same results. The developed approach bridges between physics-based building energy models and strong optimization tools available in python which can allow achieving global optimization.

Table of Contents

List of Figures

List of Tables

Acknowledgment

I want to express my thanksgiving to my God and Savior Jesus Christ for helping me in the educational journey and keeping strong and focused throughout my thesis project.

I would like to thank Dr. Hamidreza Najafi for helping me find my passion in thermal fluid science and energy analysis. He provided me with advice, experimentation tools, and funding to help achieve research goals.

I thank my sister Rosine Dusabe for the encouragement and unfailing love that kept me moving in my studies.

Dedication

I want to dedicate this thesis to my Lord and Savior, Jesus Christ, who provided me with a gift of education and enabled me to write my master's thesis.

Chapter 1
Introduction

i. Motivation

Building energy efficiency can be complex to achieve since several interconnected subsystems such as building's architectural design, envelope materials, energy end-users (e.g., HVAC systems, lighting, water heater, etc.) as well as building's operation, control, and maintenance impact the building's overall energy performance. This issue has been served well with the current simulation tools such as EnergyPlus, eQuest, DesignBuilder, and others that can model buildings as complex systems and allow users to optimize the designs to attain higher building performance. While these software tools facilitate the calculation of building energy consumption on an hourly basis for various building designs and operating schedules and essentially allow developing parametric studies for each design/operation variable, the global optimization of the building remains to be a challenging target given the numerous parameters involved in building energy analysis and their correlations.

In order to facilitate the optimization process using building energy simulation tools (such as EnergyPlus), one can use robust optimization tools available through coding platforms such as MATLAB or Python. Most studies aim to find the optimal building design envelope and system control strategies that provide minimal energy consumption, cost and better thermal comfort, and better environmental impact [1]. Even though the coupling of simulation software and optimization tools is being used extensively, this approach still has a downside since it requires significant simulation evaluations that make the whole process computationally costly, not to mention a higher possibility of errors. As a result, the approach may even become unfeasible. Some researchers have performed optimizations using both simulation software like EnergyPlus, and TRNSY and generic optimization software like GenOpt and BeOpt. Karaguzel et al. used the GenOpt software in conjunction with EnergyPlus to determine optimal thermal insulation thickness for a building envelope and glazing types [2]. Rabani et al. proposed a multi optimization approach for automating the process of determining the optimal measures that lower building energy consumption and obtain a zero energy building performance, thermal and visual comfort. The study used a coupling of an Indoor climate and energy simulation software (IDA-ICE) and the GenOpt software [3]. Corbin et al. developed a predictive model control environment that integrates Matlab and EnergyPlus to perform real-time optimization using a building automation system [4].

One method that can mitigate the computational cost associated with optimization and simulation tool coupling mentioned above is the use of simpler models, also called "surrogate models" or "meta-models." These models predict the behavior of the complex simulation models through computationally efficient approaches. One way this is achieved is by using data sets deployed to develop an artificial[1] neural network-based meta-model. This study presents an optimization of building systems and design parameters to minimize electrical energy consumption. The optimization uses a genetic algorithm applied on a surrogate model developed using a machine learning approach; furthermore, the genetic algorithm's performance is compared with that of a Bayesian optimization algorithm.

ii. Problem Statement

Although building energy simulation tools, such as EnergyPlus, are very powerful in predicting the hourly energy consumption of buildings and allow conducting various parametric studies to understand the impact of each parameter on the performance of the building, they do not facilitate the optimization process for achieving a global optimal. Also, they are considered computationally expensive, which is a major concern given the large number of variables related to the design and operation/control of a building. In order to address this challenge, the present thesis attempts to develop an effective approach that is both computationally efficient and facilitates the global optimization of the building parameters. The proposed approach consists of three stages, including developing a physics-

[1] To simplify the reading, the term "artificial" will not be used and will refer to "neural network" (NN).

based model of a building in EnergyPlus, developing a data-driven model through an artificial neural network to be used as a surrogate model and, optimizing the neural network-based model using advanced optimization algorithms available in coding tools (i.e., Genetic Algorithm in Python). The neural network-based model is trained using the physics-based model and performs substantially faster than the base model. This facilitates the computational effectiveness of the optimization process. In order to assess the performance of the proposed approach, it has been implemented on the Florida Tech Folliard Alumni Center (FALC) and discussed in detail in this thesis.

It should be noted that this study establishes a methodology for developing data-driven surrogate models with which one can optimize using available optimization packages within python or MATLAB. The methodology can be applied to different building-related solutions, such as building automation system adaptative control. Additionally, the specific design of this project can be used by building energy modelers during the ideation process before implementing more complex solutions using simulation software like EnergyPlus.

Florida Institute of Technology has completed the construction of the Folliard Alumni Center (FALC) in October 2020 (Figure 1 and Figure 2). The FALC is a zero-energy building. This project was made possible through funding from the Florida Department of Agriculture and Consumer Services and multiple industry partners. The building is an example of a high-performance building with several interrelated subsystems employed to ensure its energy efficiency and net-zero energy capabilities. This building underwent vigorous simulation and optimization process to achieve its current near-zero energy building standing. However, given the excessively high computational cost, a global optimization process was not conducted prior to the building construction. A surrogate model of the building is created using an artificial neural network and optimized with the genetic algorithm. The neural network model is trained on building simulation data obtained using a reference model of the current building simulated with EnergyPlus. Eight parameters, including; wall and roof R-values, window U-value, Window solar heat gain coefficient (SHGC), HVAC seer value, cooling setpoint, light power density, and PV tilt angle, are optimized to maximize the building's energy production through solar panels and minimize energy consumption simultaneously.

Figure 1 FALC (side view)

Figure 2 FALC (inside view)

iii. Literature Review

The so-called "artificial" neural networks, as opposed to the biological neurons in our brain, find their origins in neuroscience in the desire to understand and imitate the functioning of the human brain. In 1943, Mcculloch [5] invented the first formal neuron, modeling a biological neuron with binary behavior, showing that their model can theoretically perform complex arithmetic operations. In 1949, Hebb proposed a law of learning by observing classical conditioning in animals [6]. He explains the behavior change observed following repeated solicitations by introducing the concept of synaptic plasticity. This suggests that when two neurons are connected, the link between the mounting strengthens or creates a new one. It was not until 1958 that Rosenblatt [7] created the perceptron, a binary classifier believed to be the first modern neural network. In 1960, Widrow and Hoff [8]developed a neural network which they call ADALINE (adaptive linear neuron), using a new learning rule at the origin of the backpropagation algorithm of the gradient, widely used today. Figure 3 below shows an example of a typical neural network structure that is commonly used these days.

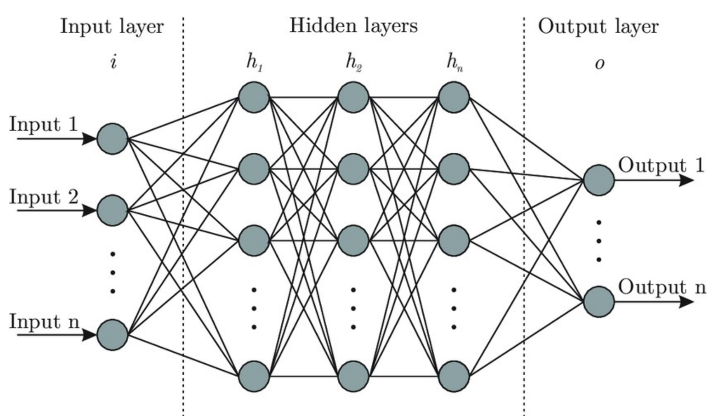

Figure 3 typical NN Architecture [9]

The applications of neural networks in the building sector are numerous and diversified. Below is an overview of where NNs are implemented in the building sector. A total of 89 published articles from 1998 to 2018 were analyzed [10], divided into four categories ranging from design to renovation of a building, as shown in Figure 4. The distribution of articles in each category is based on practical working methods. This categorization remains relatively subjective since some methods may apply to more than one category.

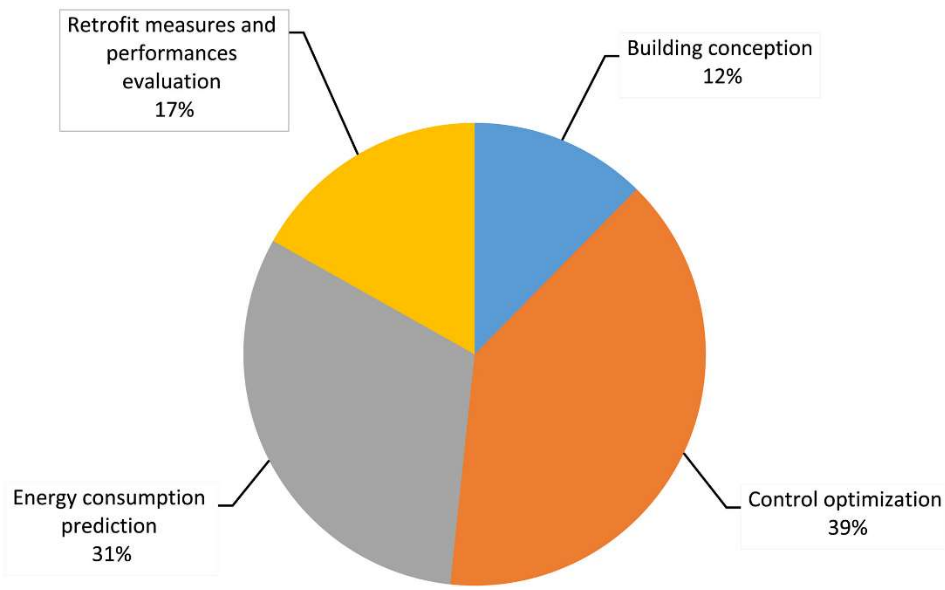

Figure 4 Distribution of the publications[2] in connection with building energy

Using prediction models such as meta-models for modeling optimization proves they are far more computationally efficient than the original simulation model that requires a long optimization process. However, developing this meta-model is a challenging process as it requires high prediction accuracy, and this is a problem that needs to be solved [11]). Research efforts in this area are considered on an optimization method that integrates neural networks (ANN) with a genetic algorithm (G.A) proposed to minimize energy consumption

[2] analysis carried out on the publications by Elsevier from 1998 to 2018 in connection with building energy

and implementation costs for residential buildings. After the ANN-based predictive model was developed, it was used for optimization; HVAC and building envelope parameters were optimized to achieve high energy efficiency [12].

Neural network-based optimization was proposed and experimented on a two-story building in Italy. The study involved applications on building energy management and indoor climate control. This study showed high energy savings and better occupant comfort compared to traditional building control systems [13].

A reinforced learning control strategy for building HVAC was developed using neural network models to reduce energy costs and demand charges. The model balances building needs with low electricity demand day times to achieve its goals, as presented by Jiang, Z. et al. [14]. Similarly, an NN-based optimizer approach is performed on a swimming pool heating system in Hong Kong. The model's objective is to maximize the thermal comfort ratio while reducing electricity consumption and lifecycle costs. The model is then optimized using a non-dominated sorting genetic algorithm for this multi-objective optimization problem; Li, Yantong, et al. [15]. Li, Hongyi, et al. performed a co-simulation between EnergyPlus and MATLAB. The study used a biased RELU neural network model to predict and optimize temperature control strategies to minimize building energy consumption [16].

Another neural network-based prediction model for annual energy consumption forecasting and thermal comfort index was developed and optimized with a multi-objective G.A. to determine the optimal design properties for a building envelope [17]. Similarly, a multi-objective optimization model that combines ANN with GA was developed by Asadi et al. to identify effective strategies for building energy retrofit that can minimize building energy consumption, implementation costs and maximize thermal comfort [18]. Bamdad, Keivan, et al. used a sampling strategy involving several surrogate models to perform building energy optimization. The given approach is compared to the traditional simulation-based optimization. Results show that the surrogate model-based optimization provided results faster and generated similar best solutions [19].

There have been significant research efforts on building performance optimization, as summarized above. The research results indicate how vital building design and operations optimization are to affect energy use positively. However, there is still a need to investigate surrogate model-based optimization specifically for zero energy buildings. These buildings are specifically designed to achieve maximum efficiency, thus undergo rigorous modeling before construction and operation. The surrogate model approach can produce models with good accuracy and less computational time. This study presents the optimization of a feedforward neural network-based surrogate model using a genetic optimization algorithm. The performance of the genetic optimization is then compared with one done using the Bayesian optimization. Building design and control parameters are optimized to minimize total building electricity consumption. Two surrogate models are developed that represent the performance of a zero energy building during regular operations and unforeseen pandemic operations.

iv. Methodology

The present research proposes an efficient approach for modeling and optimizing building energy performance through a three-stage methodology implemented on FALC as a case study and described below. Figure 5 outlines the project implementation flowchart.

Stage 1: Develop a physics-based model of the building using EnergyPlus. An input data file (IDF) is developed containing simulation data used by EnergyPlus. The reference idf is edited to change building design parameters[3]. A Python Eppy package was used to automate the entire process of generating input files (.idf) and reading output files (.eso) from EnergyPlus associated simulations. The model is edited and simulated multiple times to generate a dataset containing design parameters and corresponding annual energy values. Design parameters were defined with ranges of minimum-maximum values, and the dataset was uniformly distributed.

[3] Idf: input data file, edited using Python package called Eppy

Stage 2: Develop a surrogate model using artificial neural networks. The generated dataset is loaded to the predictive model to train it to predict energy consumption, production, and net site energy.

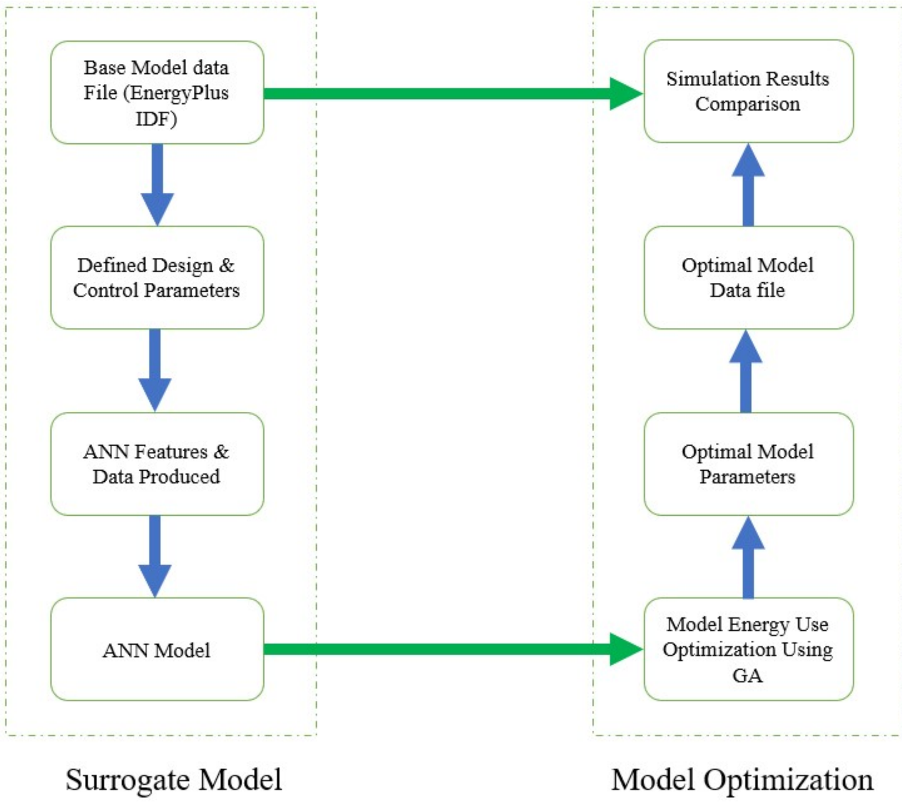

Figure 5 ANN Surrogate Model Optimization Framework

The NN model takes the sampled inputs and the associated output values (machine learning features).

Stage 3: Optimization of the surrogate model. The neural network model is then used as a black-box for the Genetic optimization algorithm. The black box model is a scoring function to estimate building energy consumption. Additionally, a Bayesian Optimization was tested to compare the optimization performance of the two.

Chapter 2
Building Energy Modeling

i. EnergyPlus Simulation and Results

A building energy model for the Folliard center (FALC) was developed first, with a 3D model (Figure 6) developed using OpenStudio's SketchUp plugin. This model was then further developed using OpenStudio's application to add all connected systems, including specific heat pumps, water heating system, lighting system, PV system, and building envelope materials.

The baseline model for this investigation was an EnergyPlus medium-sized office building (Figure 6) designed to operate as a zero energy building. This is a one-floor building with an east-facing orientation. There are 5 office rooms, 1 conference room, making up 3 core thermal zones. The entire floor space is 3,500 ft^2 (334 m^2). The conference room height is 16.2ft, while the rest of the building is at an overall height of 14.8 ft, with a glazing ratio of 7%.

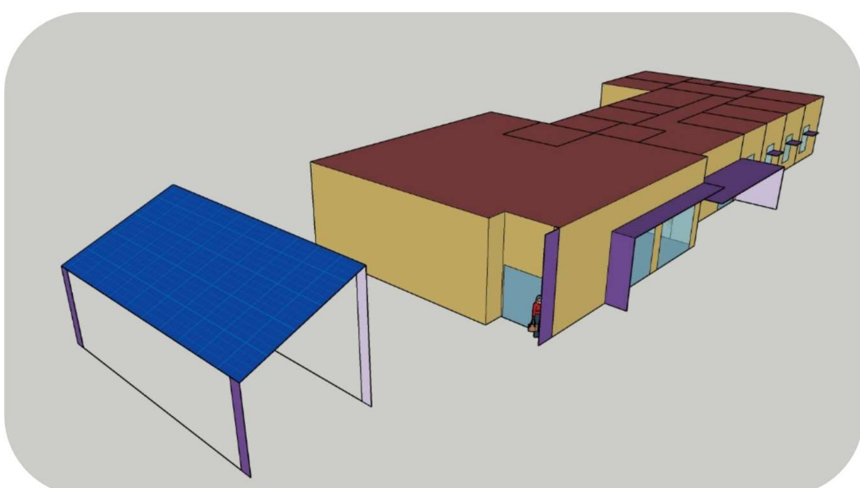

Figure 6: 3D model

Although the FALC reference building's original geometry was preserved, the envelope materials and their thermophysical characteristics, PV system orientations, and HVAC system efficiencies were changed to generate different building designs used in training the NN. The building is located in Melbourne, Florida, which has a hot and humid climate, and simulations were performed using the Melbourne International Aiport weather TMY3 file [20]. The reference model cooling and heating setpoints are 22°C and 18°C, respectively, with setback temperatures of 24°C and 16°C.

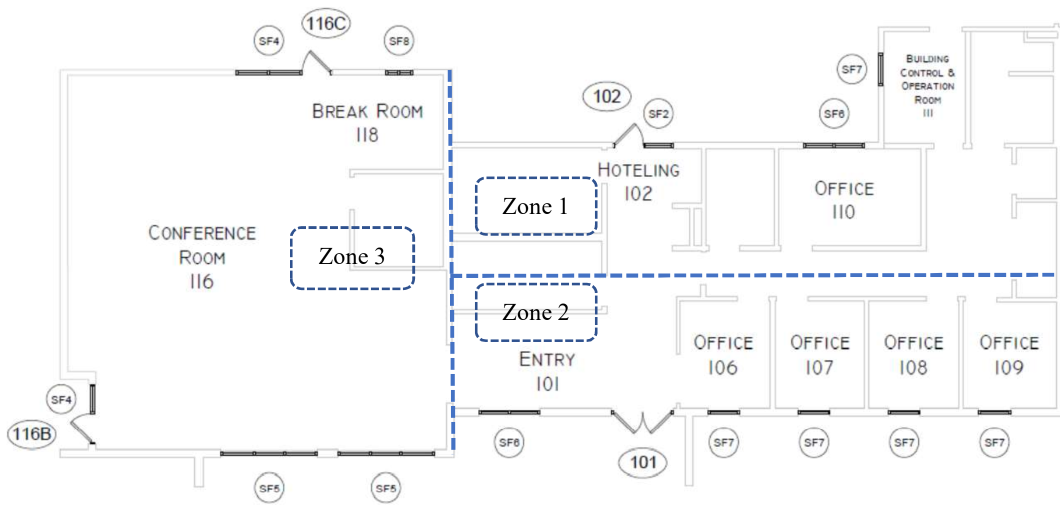

Figure 7 FALC building layout

3D modeling is the first step that most energy modelers do when they start building energy modeling. Several Graphical User Interface (GUI) enhanced software tools such as those in OpenStudio enable users to easily add building systems and their criteria before EnergyPlus can perform the actual simulation. However, for this project, all systems were modeled individually using EnergyPlus since the Eppy Python package can only communicate with the EnergyPlus API. Furthermore, when working with OpenStudio and exporting the EnergyPlus IDF file, it seems not to bring all modeled systems with the exported file.

Figure 7 shows the FALC building layout used to develop a SketchUp 3D model that was essential the backbone of all building simulations. The building is divided into three thermal zones, as indicated by the blue dotted lines on the layout.

Table 1 Folliard center thermal zones parameters

Thermal Zone	*Zone 1*	*Zone 2*	*Zone 3*
Cooling COP	3.7	4.4	3.9
Heat COP	2.9	2.9	2.9

Table 2 HVAC system specifications

Condensing Unit	*CU-1*	*CU-2*	*CU-3*
Compressors	1	1	*1*
Nominal Tons	3.0	2.5	*4.0*
Suction Temperature	45	45	*45*
Ambient	95	95	*95*
Manufacturer	TRANE	TRANE	*TRANE*
Model Number	4TWR7036	4TWR6030	*4TWR7048*
S.E.E.R.	17.5	17.0	*17.0*
Voltage/phase	208/1	208/1	*208/1*
MCA (AMPS)	21.0	17.0	*28.0*
MOP (AMPS)	35.0	25.0	*45.0*
Heating Capacity at 47 °C (MBH)	33.2	28.8	*46.0*
COP at 47 °C	3.9	4.4	*3.7*
Air Handling Unit	**AHU-1**	**AHU-2**	***AHU-3***
Manufacturer	Trane	Trane	*Trane*
Model Number	TAM9A0C36 V31	TAM9A0B30 V31	*TAM9A0C48V 41DA*
Cabinet	Vertical	Vertical	*Horizontal*
Cooling Coil	Single Circuit	Single Circuit	*Single circuit*
Electric Heater (kW)	5.76	5.76	*7.2*
Total Cooling	35,700	30,000	*48,000*

Capacity (BTU/hr)			
Sensible Cooling Capacity (BTU/hr)	27,400	22,400	*36,800*
Fan Motor Size (HP)	1/2	1/2	*3/4*
Air flow (cfm)	*35*	*30*	*95*

The three thermal zones are served by individual heat pumps, each having its condenser unit (CU) and air handling unit (AHU). Table 1 shows the SEER values for each Heat pump used. Additional HVAC system specifications are listed in Table 2.

Additionally, the photovoltaic (PV) solar system was modeled to determine onsite power generation that could be achieved. Since this study involves training the NN with different PV system tilt angles to predict the power production at each tilt angle, thus the modeled PV panel system in EnergyPlus provides the training data for the PV system. The PV system is made of a canopy with solar 30 solar panels. The system has a 9.4kW power capacity and can generate a minimum of 11MWh annually. Table 3 shows the FALC system specification that is used in the building energy model.

Table 3 PV system specifications

PV branch	***Branch 1***	***Branch 2***	***Branch 3***
Power Capacity	3000	3000	3000
Panel Modules	10	10	10
Inverter efficiency	97.5%	97.5%	97.5%
Tilt Angle	20^{o}	20^{o}	20^{o}

Other modeled systems that significantly impact the overall building's energy consumption include lighting systems, water heating systems, and building envelopes. Since this study aims to produce a virtual model that can be optimized for selecting the best building design parameters, the reference building energy model used to train the NN closely reflects the current building design parameters implemented on the FALC.

Table 4 FALC building envelope specifications

Area	***Wall layers***	***Overall Wall R-Value***
Wall	8” CM block wall with 3- part stucco (layers listed from the outside to the inside of the wall): 3-part stucco, painted 8” CMU block (hollow) 1 ½” foil-faced rigid insulation (MIN. R-5.5 per inch) R-8.25 for 1.5”. 7/8” metal Furring channel at 16” O.C. 5/8” Gypsum Board	*R-10.1*
Roof	*TPO Roofing System with 3 layers (layers listed from the outside to the inside of the wall):* *1. Mechanically fastened TPO roofing system (R-0.24)* *2. 6” Polyiso insulation (R-30)* *3. 5/8” Plywood decking (R-1.25)*	*R-31.49*

Table 4, seen above, represents the envelope layers and their respective materials that enclose the building. These materials were modeled in EnergyPlus to reflect the actual building envelope as implemented at FALC. Eventually, the wall R-values are changed to depict using different insulation. The different wall designs and their impact on the overall building energy consumption and other design parameters are used to train the NN surrogate model. Furthermore, the window types and their heat transfer coefficients and solar heat gain coefficients, U-value and SHGC, used at the FALC are also modeled in the reference model. Table 5, shown below, lists U-values and SHGC for the FALC windows.

Table 5 Windows thermal performance

FENESTRATION	MAX U-FACTOR (FIXED/OPER)	MAX SHGC (ALL AXIS)
INSULATED GLASS, LOW-E	1.22	0.40

Additional BEM design inputs include the lighting controls, cooling, and heating setpoints performed based on occupancy schedules. The occupancy schedule is based on the school calendar for office work, ideally followed for the office occupants at FALC and a 3-hour weekly full occupancy of the conference room at FALC.

Table 6 Glass types in the FALC building

Glass Type	***Thickness (in)***	***Thickness (mm)***	***Name***
GL1	1.3125	33.34	*Clear Solar Ban 60 #2*
GL2	0.5625	14.2875	*Clear Solar ban 60 #2*
GL3	*0.25*	*6.35*	*Clear Tempered*

The building is made of 13 windows that are of 3 different types, shown in Table 6. For each type of window, the thickness and names are provided. It should be noted that the windows are Low-E double glazed; however, coatings and thickness sizes are different.

Table 7 Windows reference table

Reference	***Glass Type(s)***	***Door/Window***	***Quantity***
SF1	GL2	Main Door	*1*
SF2	GL1 & GL2	Door (GL2) & Window (GL1)	*1*
SF3	GL1 & GL2	Door (GL2) & Window (GL1)	*1*
SF4	GL1 & GL2	Door (GL2) & Window (GL1)	*2*
SF5	GL1	Window	*2*
SF6	GL1	Window	*2*
SF7	GL1	Window	*5*
SF8	GL1	Window	*1*
SF9	*GL3*	*Window*	*1*

Figure 7 references different windows locations, and their reference names are listed in Table 7, along with the number of available windows of each type.

Table 8 Light fixtures used at FALC

Lighting Fixture Schedule						
#	TYPE	MANUFACTURER	MODEL	LAMP	LUMENS	W
B	Recessed LED 2X2	CREE	CR22-32L-ACK-POE	LED	3,200	32

Lights are also included as inputs to the EnergyPlus model and their corresponding power densities for each building area. The lights selected provide the best energy savings since they are all LEDs and can be automatically controlled using a building automation system to achieve even more savings from scheduled operations.

Table 9 Interior Lighting Power Density

	Allowance			**Luminaries**				
Area	**Area (ft^2)**	**Allowed (W/ft^2)**	**Allowed Watts**	**#**	**Type**	**Watt**	**Designed Watts**	**Designed (W/ft^2)**
Conference Room	1,282	1.2	1,538	31	8	32	992	0.8
Kitchen	98	1.1	108	2	8	32	64	0.7
Storage	89	0.8	71	2	8	32	64	0.7
Woman R/R	127	1.0	127	2	8	32	64	0.5
Man R/R	107	1.0	107	2	8	32	64	0.6
Computer Area	55	1.1	61	1	8	32	32	0.6
Entrance Hall & Other	327	1.0	327	6	8	32	192	0.6
Hallway	250	0.8	200	5	8	32	160	0.6
Mech. Room	72	0.8	58	1	8	32	32	0.4
Office 1	148	1.1	163	3	8	32	96	0.6
Office 2	107	1.1	118	2	8	32	64	0.6
Office 3	107	1.1	118	2	8	32	64	0.6
Office 4	105	1.1	116	2	8	32	64	0.6
Office 5	110	1.1	121	2	8	32	64	0.6
Office 6	101	1.1	111	2	8	32	64	0.6
Mix R/R	103	1.0	103	2	8	32	64	0.6

Janitor	23	0.8	18	1	8	32	32	1.4
Storage	41	0.8	33	1	8	32	32	0.8
Total	3,252		3,497			2,208	2,208	0.7

Additional lights are modeled to represent the designed exterior lighting and their power densities. Exterior light specifications are shown in Table 10.

Table 10 Exterior Lighting Power Density

	Allowance			**Design**			
Area	**Linear (ft)**	**Power (W/ft)**	**Total**	**#**	**Typ**	**Watt**	**Total**
Main entrance & other doors	310	0.7	217	9	W	13	117
Base allowance	-	-	600				
		Total	817			Total	117

The Lutron Roller 100 Shade has the following specifications:

- Ultra-quiet operation:
- Shades move in perfect unison and exact alignment within 0.125 in (3 mm) accuracy at all times.
- Smooth, silent starts and stops.
- Programmable stop points.
- It provides maximum window coverage with small light gaps.
- Gaps are symmetrical on both sides of the shade.
- Easy-to-read and easy-to-use controls.
- Operating voltage: 24-36 V low-voltage power.

The Shade fabric has the following specifications also presented in Table 11

- The shade is a Lutron family name of the shading fabric: “Sheerweave 4900” (Model number P06-49- 3).
- Thickness: 0.0270 in (0.7 mm).

- Material: 17% Polyester, 83% Vinyl on Polyester.

Table 11 shade fabric energy specification

Openness Factor (OF)	Visible light Transmittance (Tv)	Solar Transmittance (Ts)	Solar Absorptance (As)	Solar Reflectance (Rs)
3%	10%	11%	8%	81%

ii. BEM simulation results: regular building operations

This section presents simulation results for the modeling representing the normal building operations initially designed for the building. At normal operations, the HVAC system is scheduled to operate only during occupancy, including warm-up times. The HVAC system is off during unoccupied hours, including night times, weekends, and holidays. Additionally, the conference room is scheduled to operate 3 hours a week, with which the HVAC for the conference room is fully operating, and the lights are turned on.

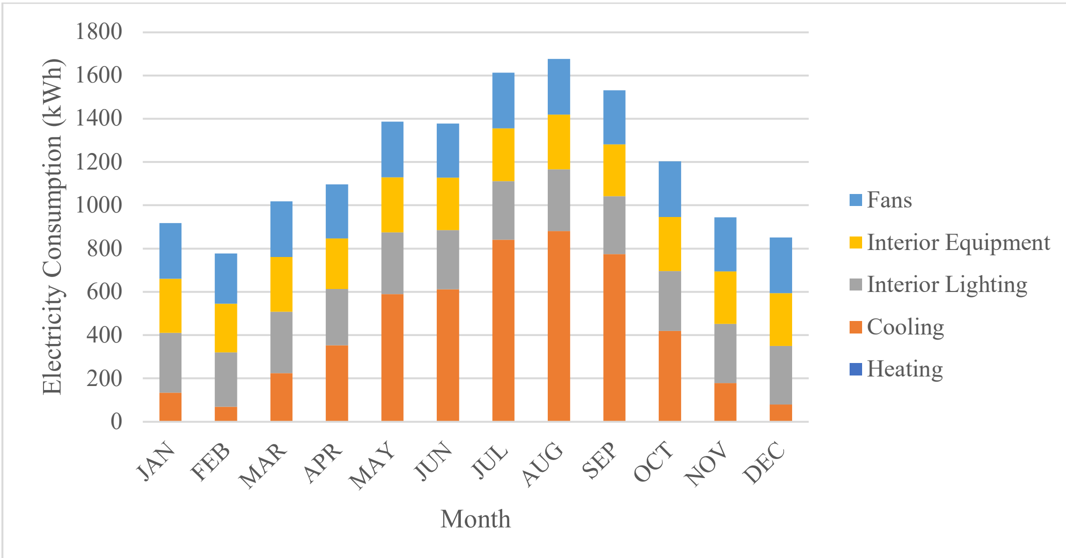

Figure 8 BEM prediction for regular operating schedule: monthly energy consumption

Figure 8 is a representation of the electricity consumption by all major end uses in the FALC. These are results obtained from the building energy simulation performed in EnergyPlus. Results are relatively close to the initial building energy analysis done before the construction of the FALC. As expected, most electricity is consumed for cooling down the building. Space cooling accounts for about 36% of the total electricity consumption, as shown in Figure 9. Interior lighting and fans also consume about 23% and 21% of the total consumption, respectively.

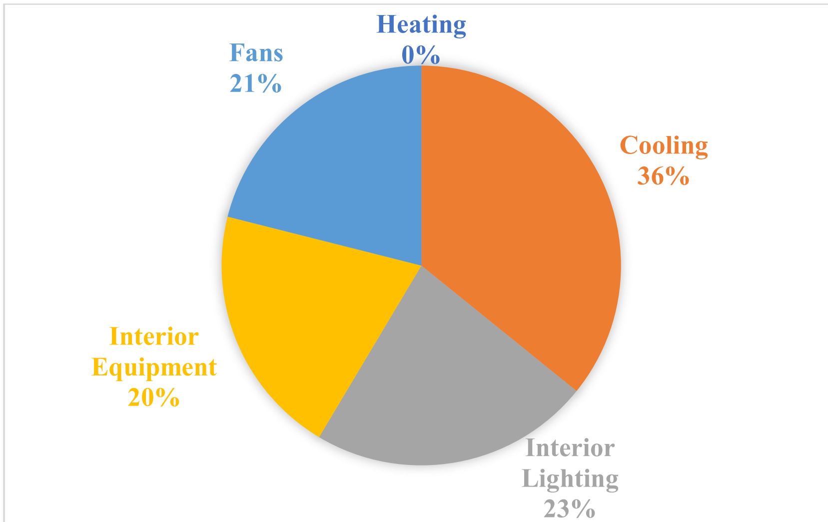

Figure 9 Regular BEM: end-use Annual energy share

After performing the simulation and cross-checking with the initial design simulations to confirm the degree of accuracy, the BEM model was ready to be used for NN training. The simulation results presented agree with simulations performed with other modeling software taking into account the 3 hours weekly schedule for occupancy in the conference room. However, it should be noted that the actual energy consumption of the FALC for the past year, 2020-2021, is relatively higher than the simulation results in this study. The cause of the discrepancy arises from the full capacity operation and using the conference room as a classroom throughout the day. Thus the weekly schedule is no longer a 3 hours operation for the conference room. On top of the changed schedule, the BEM does not include the EV charging station, which has a tremendous impact on the overall energy consumption. The EV station is currently used throughout the day to charge two electric vehicles.

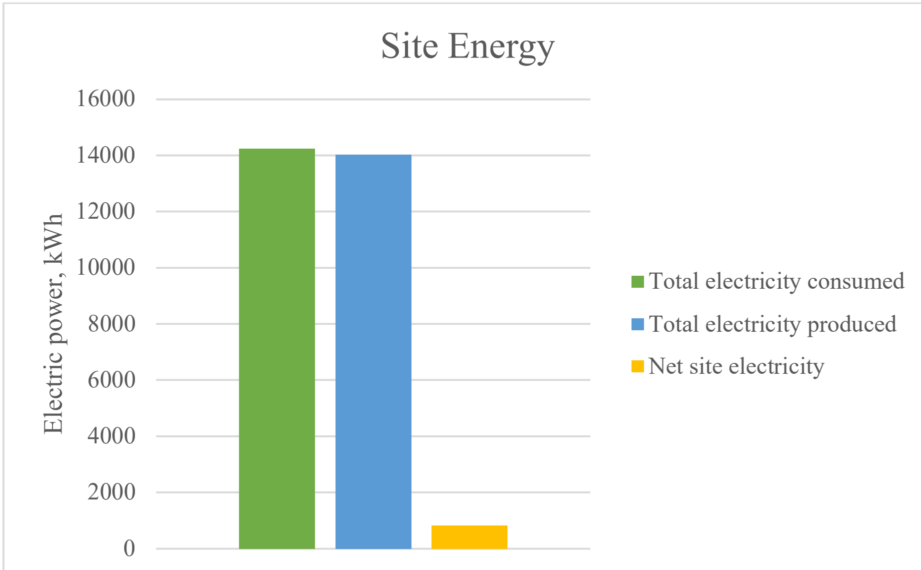

Figure 10: Regular BEM: annual simulation results

The EnergyPlus simulation is set to only output 3 annual values, as shown in Figure 10. These results are the ones needed in further surrogate model development.

Table 12 Summarized annual energy use

Electricity consumption	14,243 kWh
Total annual electricity peak demand	10.15 kW
EUI (Energy Use Intensity)	15 kBtu/ft^2/year

A condensed form of energy results is only used in the development of the NN-based surrogate model. The data of interest are the annual electricity consumption by the facility, production by the PV system, and the net site electricity (difference between electricity produced by PV and electricity consumed by building) for subsequent 8 design inputs to the

BEM. The net site energy is significant because it determines the electricity cost for imported electricity or the electricity sold in export to the grid. Figure 10 shows example data of what is extracted from the results after each simulation run.

iii. BEM Simulation Results: Operations during the pandemic

This section presents simulation results for the modeling representing the pandemic building operations. During the pandemic operations HVAC system is scheduled to operate continuously during occupied and non-occupied hours. The schedule ensures that there would be maximum fresh air circulation to prevent COVID-19 from spreading through the HVAC system. Additionally, the conference room is scheduled to operate more hours per week, similar to the office activity schedule of 8 a.m to 5 pm; the conference room is used as a classroom throughout the week.

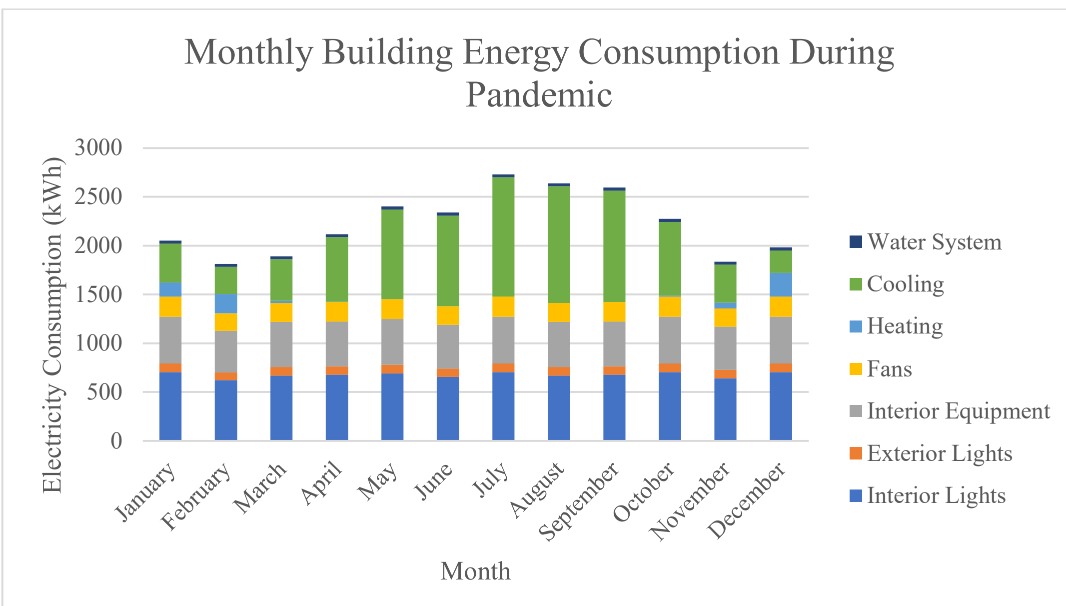

Figure 11 Monthly energy results from a pandemic simulation modeling

Figure 11 represents the electricity consumption by all major end uses in the FALC during the pandemic times. These are results obtained from the building energy simulation performed in EnergyPlus. Results are relatively close to the actual building energy consumption. As expected, most electricity is consumed for cooling down the building.

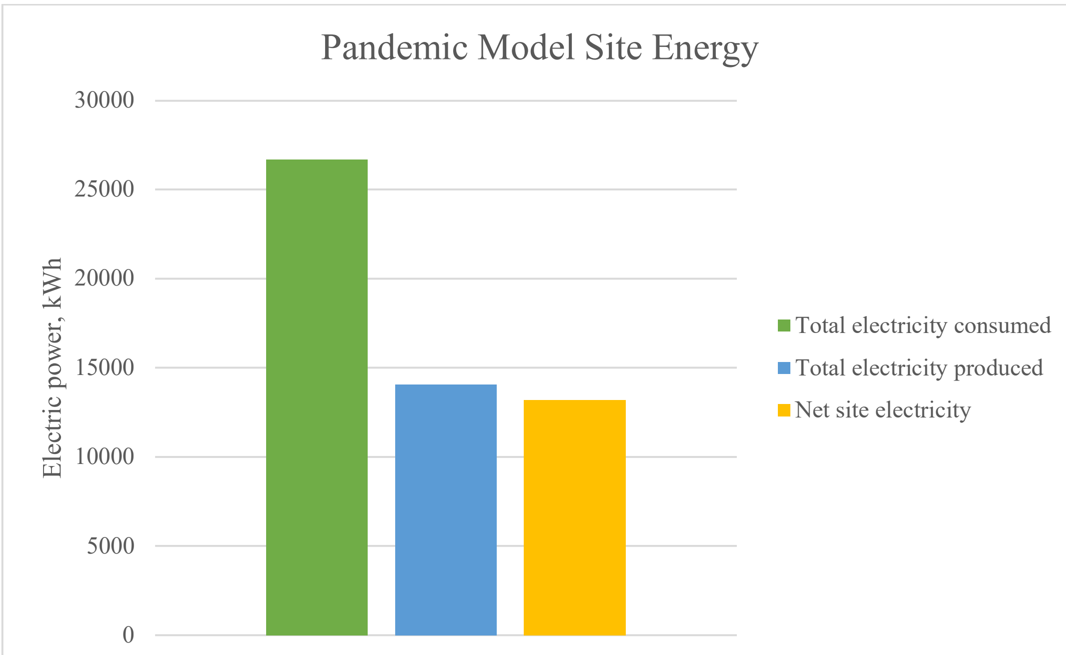

Figure 12 Pandemic annual simulation results

The EnergyPlus simulation, in this case, is also set to only output 3 annual values, as shown in Figure 12. These results are the ones needed in further surrogate model development.

Table 13 Summarized annual energy use during pandemic

Electricity consumption	26,674 kWh
Total Annual electricity peak demand	20 kW
EUI (Energy Use Intensity)	25 kBtu/ft^2/year

iv. Automated EnergyPlus Simulation

Performing multiple simulations can be made easier by automating the process. In this study, Python is being used to perform simulations on a locally installed EnergyPlus software iteratively. Thus, running multiple simulations is automated to reduce the time that would otherwise be taken to manually run individual simulations and later gather all data needed to develop the NN surrogate model. A Python package, Eppy, was used to work with EnergyPlus using Python.

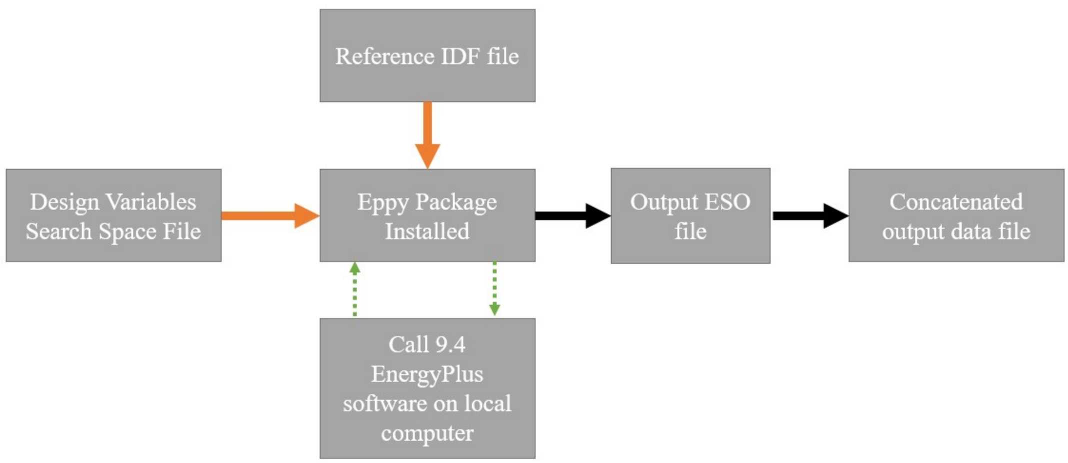

Figure 13 Auto-simulation Program Framework

Figure 13 represents the framework of the co-simulation program developed to allow data generation. A reference IDF file containing all design variables and components of the building energy model is loaded into the Python program. The reference IDF file and the variable search space file are inputs to the Eppy API that runs the EnergyPlus on a local computer. IDF loading and editing are performed at the fourth stage (3) of surrogate model development, as illustrated in the methodology section of this study.

Table 14 design and control parameters

Parameter	Minimum	Maximum
Wall R-value	8	25
Roof R-value	25	45
Window U-value	1	7
Window SHGC	0.2	0.7
Cooling Setpoint	21	26
HVAC Cooling COP	2	5.5
Light power density	2	15
PV Tilt Angle	10	35

Table 14 shows all 8 design and control parameters changed at each simulation run to generate subsequent energy consumption, production, and net site energy. A total of 200 changes are performed for each design parameter, and values are randomly selected between the range provided in table 5 for each variable. There are 200 rows and 11 columns containing design variable values and simulation results represented as annual electricity consumption, production, and net site energy. Each row of data has its specific IDF file containing the design variables of that row. This IDF can be simulated in EnergyPlus to produce the desired output; only 3 annual results are used for this study.

At this stage, the Python package has helped create a database that can be used for the NN model development. Suppose a random user wants to perform a test using the Python auto-simulation program developed here. In that case, they need a reference IDF file that a user wants to study, and specified ranges of the design variables desired to change within the IDF. Thus, the auto-simulation program can quickly generate a dataset for NN model development, as the case for this study, or other parametric analysis tools desired by the user. The auto-simulation program makes data generation easier and computationally cheap.

```python
def get_mapped_vals(col):
    if col=='Wall R value':
        return 'Thermal_Resistance'

    elif col=='Roof R value':
        return 'Thermal_Resistance'

    elif col== 'HVAC COP':
        return 'Gross_Rated_Cooling_COP'

    elif col=='Window U value':
        return 'UFactor'

    elif col== 'Window SGHC':
        return 'Solar_Heat_Gain_Coefficient'

    elif col== 'Cool setpoint':
        return 'Value_Until_Time_1'

    elif col=='Light Pwr Density':
        return 'Watts_per_Zone_Floor_Area'

    elif col== 'PV Tilt angle':
        return 'Tilt_Angle'

def make_mappings(obj_num_df, obj_name_df):
    mapping= dict()

    for col in obj_num_df.columns:
        mapping[col]= [obj_name_df[col].dropna().tolist(),(obj_num_df[col]
.dropna().astype(int)-1).tolist(),get_mapped_vals(col)]
    return mapping

mappings= make_mappings(obj_num_df,obj_name_df)
```

Figure 14 Python auto-simulation snippet

An illustration Python script used to map to the design variables in the reference IDF file is shown in Figure 14. Once the design variables are mapped, then the program can change their values and save a new IDF file with the newly changed values of each mapped design variable. Table 18 in the appendix contains a database that was generated by the python auto-simulation program.

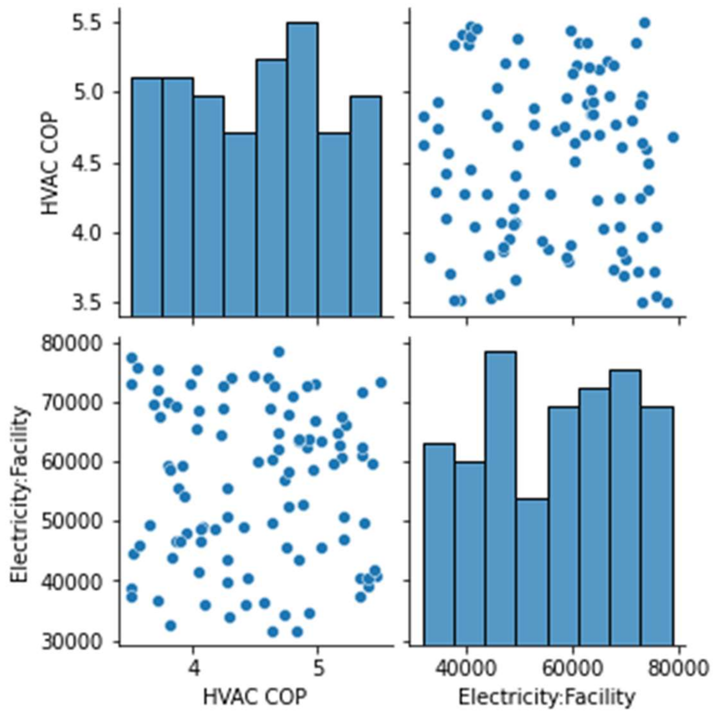

Figure 15 HVAC COP and electricity consumption Distribution

The Python auto-simulation can also help to visually inspect the data distribution and find its relation to the variable of interest. Some data may prove to have no relation to an output of interest and may not help in NN development if the NN is a linear model. Figure 15 shows one example of a plotted data distribution for design variables relative to the annual facility electricity consumption. The data distribution shown pertains to the 100 by 11 matrix of data used in this study. Figure 15 shows HVAC COP to have a poor correlation with the electricity consumed. It doesn't render it useless for NN training since the NN model developed in this study is a non-linear and multivariable prediction model and performs well with inputs that show no correlation.

Chapter 3
Neural Network Model Development

i. Data and Machine Learning

For the past few decades, data has become an essential resource for the functioning of our modern societies. Data allows the connection between artificial objects and humans to be in unprecedented ways. Each machine or human being associated with a connected object can now produce and consume data, whether or not they are aware of it. Since the start of the Internet for the general public in the 1990s, the amount of data in circulation has grown exponentially; for example, the mass of data created and copied increases by a factor of nine every five years [21].

At the same time, the progress made in machine learning, a field of study in artificial intelligence (AI), has opened up new possibilities for the use of big data. These techniques allow computers to use data to solve problems they have not explicitly programmed for based on mathematical models. The current development of machine learning techniques is inseparable from big data, which fuels these algorithms.

The desire for valuation arises the need for the veracity of the data. Indeed, the quality of analysis is impacted by the quality of the initial data. In fact, without current machine learning techniques, data would only be a shapeless mass complicated to value. Conversely, without these precious data, machine learning techniques would not have proved their effectiveness. They would have remained in the object stage study, unable to prove their value in concrete applications. The use of big data through machine learning techniques has shown its interest in various fields such as scientific research, health, transport, energy, or the economy [22]. These techniques appear in particular to reduce energy consumption in a multitude of fields of application. Thus, with a techno-centered vision often implemented, the use of big data via machine learning is gradually becoming inseparable from the fight against climate change, including the building sector, due to its high energy consumption.

ii. Artificial Neural Networks (ANN)

The human brain inspires the structure of ANN, and they depict how biological neurons signal to one another. ANN allows computer programs to recognize patterns and solve everyday engineering problems in AI, machine learning, and deep learning. In this study, they are used to model building energy consumption, production, and net site energy of a particular building.

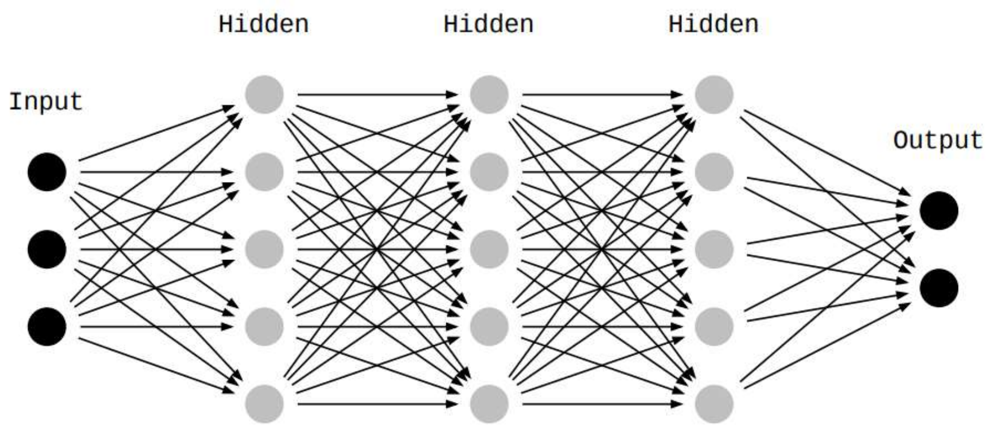

Figure 16 Example NN Structure

ANNs are made up of layers, as seen in Figure 16; an input layer, hidden layers, and an output layer. These layers are made up of interconnected neurons. Each neuron has an associated weight and threshold. An individual neuron is activated if its output is above the specified threshold value; this sends data to the next layer. Otherwise, data is not passed along to the next network layer. Different activation functions are used at each layer.

During NN model developments, a dedicated training dataset is used by neural networks to learn and increase their accuracy over time [23]. However, once these learning algorithms have been fine-tuned for accuracy, they become fundamental tools in computer science and AI, allowing users to cluster and categorize data quickly. Compared to manual identification by human specialists, speech recognition or picture recognition tasks can take minutes rather than hours. Google's search engine is one of the most well-known systems that use deep neural networks in it’s architecture.

Individual neurons are like linear regression models, with input data, weights, a bias (or threshold), and an output. An example formula at the neuron level is shown below

$$a = \sum_{i=1}^{m} (W_i X_i + b)$$

a: sum of weighted average

b: bias

w: weight

i: index

An individual neuron computes the weighted averages of its input, and the sum is passed to an activation function which produces an output to pass along to the next layer. The output of one neuron becomes the input of the next neuron. Common activation functions include sigmoid, exponential Linear Unit, rectified exponential Linear unit (ReLU), and many others. Suppose the sum of weights is activated by a sigmoid function; it would look similar to the function below.

$$output = f(x) = \frac{1}{1 + e^{-a}}$$

After determining the input layer, weights are assigned, and they help determine the importance of any input variable. Thus larger input variables contribute more significantly to the output than other inputs. The process of passing data from one layer to the next is what makes it a feedforward NN.

A more practical use for neural networks such as image recognition and other classification problems leverages supervised learning, or labeled datasets, to train the NN models. Training a model involves evaluating its accuracy, which is done by defining a cost (or loss) function.

In most regression problems, the cost function is the mean squared error (MSE). The formula for computing MSE is shown below.

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

i: sample index

$\hat{y}$: predicted outcome, y: actual value, m: total samples.

Finally, one needs to minimize the cost function to guarantee that each given observation is correctly fitted. The model uses the cost function and reinforcement learning to change its weights and bias for convergence or reaching the local minimum. Gradient descent is the method through which the algorithm modifies its weights, allowing the model to discover the best path to minimize the error represented by minimizing the cost function. The model's parameters adjust with each training case, eventually converging at the minimum.

The majority of deep neural networks are feedforward, as they flow from input to output. Backpropagation, which moves in a reverse direction, from output to input, is another way to train NN models. Backpropagation is used to compute and assign errors to each neuron, enabling users to correctly alter and fit the model's parameters [24]. Similarly, this study uses a feedforward neural network with backpropagation. The model is not developed from scratch as the Keras Python library already has several neural networks developed for users who wish to apply them to their studies.

iii. Network Development

In this study, inputs to the NN are referred to as features, and the target variables are labels. There are a total of 8 features and 3 outputs, and each feature has 100 data points. Thus, the input layers and output layers contain 8 and 3 neurons, respectively. Data preprocessing is also included during NN model development, but since this study involved automated data generation, the python co-simulation program was also used to do data preprocessing before network development. However, before data was fed to the neural network, it was standardized to have the mean of the values be zero and their standard deviation be one.

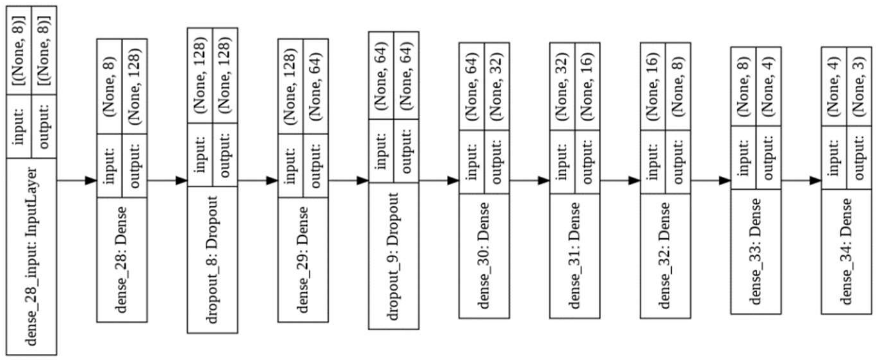

Figure 17 NN model architecture

Figure 17 above illustrates the flow of the NN model that is used in this study. It contains 10 layers, as listed in Table 15. Each layer uses the rectified linear unit, ReLU activation function; however, the output layer has no activation function. Other activation functions such as sigmoid, linear and exponential linear unit, ELU, were tested during model development; however, they were not providing model convergence. The dropout layers are used for model regularization, which helps prevent model overfitting and avoid adaptation when the neurons extract the same hidden feature from the model inputs, thus leading to overfitting [25]. The fraction of dropout (dropout rate) determines a fraction layer neurons that are shut down. This study has a dropout rate of 0.2 since the model used doesn't involve many neurons compared to other complex models with hundreds and thousands of neurons in hidden layers. The loss function is used to determine the performance of the surrogate model is the root mean squared error (RMSE).

It should be noted that the six layers were manually selected since they provide the best convergence. The development started with fewer layers and neurons. However, the model was underfitting; thus, more layers and neurons were added to prevent model underfitting.

Table 15 Feed Forward NN Model Summary

Layer number	Neurons	Activation function
1: input layer	8	ReLU
2	dropout (0.2)	
3	128	ReLU
4	dropout (0.2)	
5	64	ReLU
6	32	ReLU
7	16	ReLU
8	8	ReLU
9	4	ReLU
10: output layer	3	

The study involved developing two neural network models representing two surrogate models for pandemic and regular building operations. These models have similar architecture presented above. The pandemic BEM model was obtained by modifying the IDF design variables slightly to have a different design from the regular operations BEM model. The pandemic model has an overall building electricity consumption of close to 26.7 MWh per year and around 14 MWh of electricity produced by the PV system. This pandemic model has different building design settings relating to schedules of operations. For example, the HVAC system for each of the 3 thermal zones is set to continuous operation while the regular BEM has set schedules for the HVAC systems. However, the python co-simulation program uses the same design parameter search space for the two models when a test is performed to generate new building designs to train the neural network.

The reference BEMs have design variables assigned different numbers even though different objects share them. For example, in the reference model, there are different power densities in different building areas. However, newly generated models using the co-simulation program assign the same light power densities for all areas in the building. Similarly, the program assigns the same window value to all the windows even though some windows have different U values. As a result, data generated from the co-simulation may not provide similar energy consumptions as the original trained model since the trained model is more specific in assigning values to design variables for individual building components. The lowest energy-consuming design model generated by the co-simulation is still higher than the baseline model used for data generation.

iv. Surrogate Model for Building Operation During Pandemic

The generated data representing design parameters and their respective annual energy values were divided into a training dataset that accounts for 80% and a validation dataset that accounts for 20% of the total training data. Initially, 100 different designs models were generated for NN training and testing. However, these were not enough in producing a well-trained NN model. Additional data were generated using the co-simulation program to make 200 designs used in NN training and testing. Figure 18 and Figure 17 present the annual energy use across all 200 generated designs.

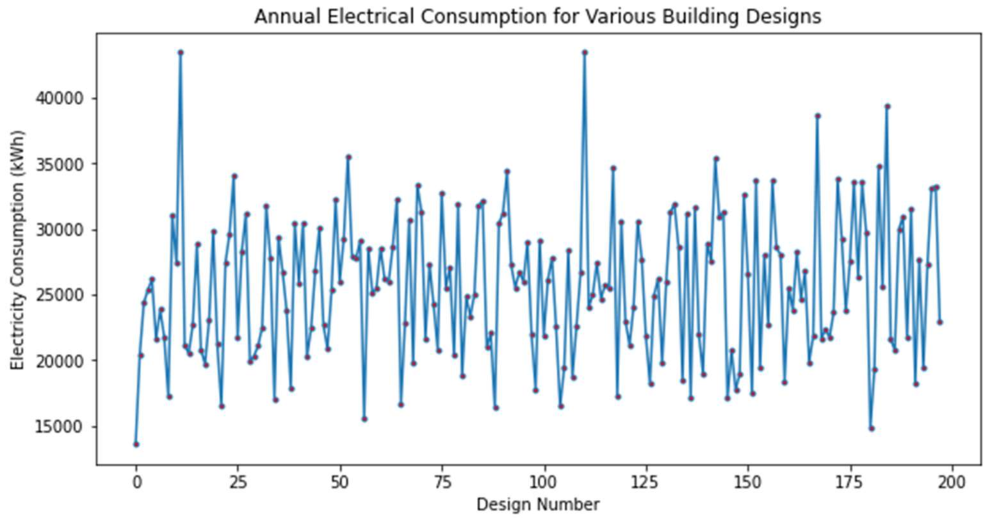

Figure 18 Energy consumptions for generated data

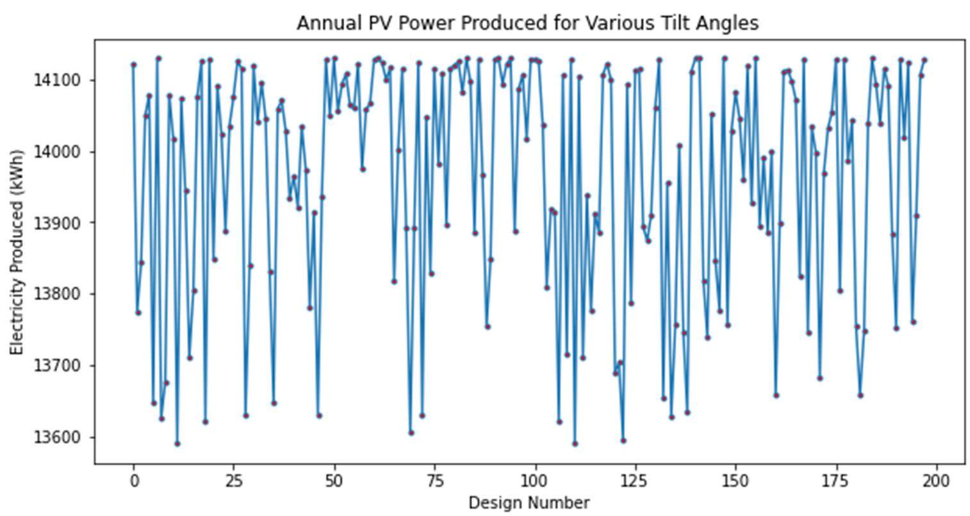

Figure 19 PV energy for generated data

The plot in Figure 20 shows the progress of the model across the epochs. 1000 epochs were used for this model. And the model achieves convergence after 400 epochs approximately. More epochs were used since fewer epochs provided very high loss values. On the training, the model eventually converges to an RMSE loss of 0.151, which is a good performance. However, this is only the first model version with minimal tuning implemented. Further tuning is done to have a well-generalized model; this is evaluated using a reserved test set to determine how well the model represents the FALC under pandemic operations.

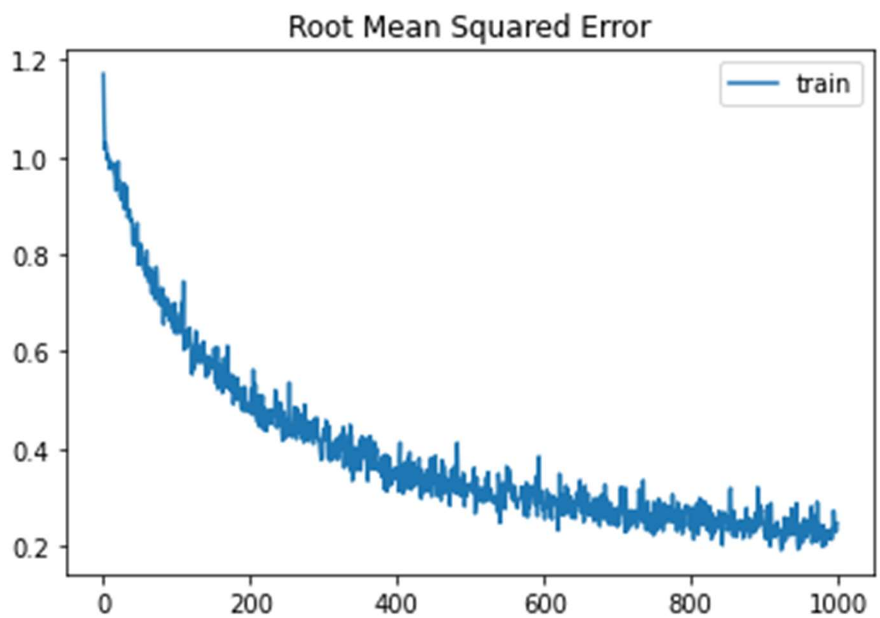

Figure 20 NN model Convergence

Plotting actual and predicted results could help visualize how the model is performing. Figure 21 shows how the predicted electricity consumption compares with the simulation data obtained from the EnergyPlus BEM on the train set. The model does very well on prediction for the training dataset. However, it should be validated to confirm its performance with unseen data.

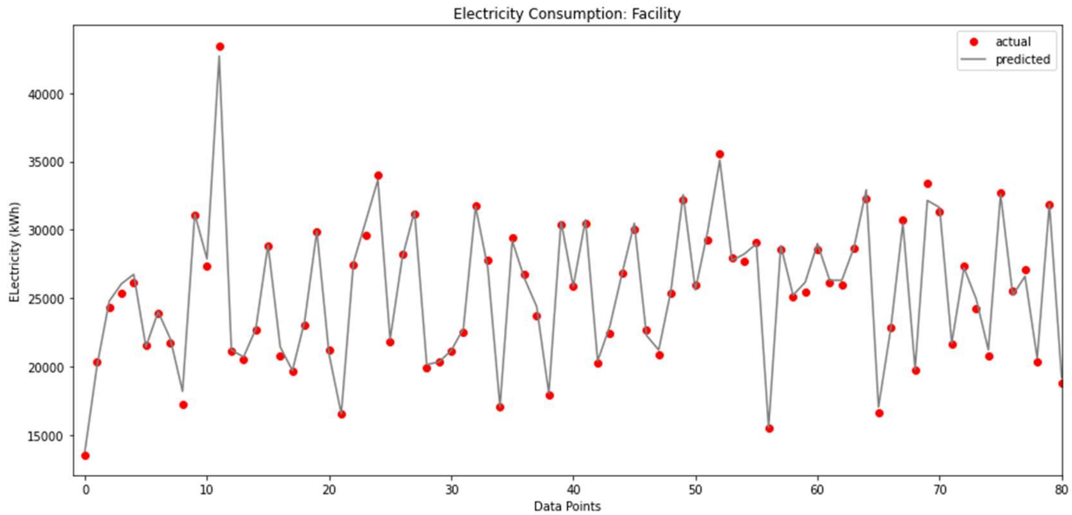

Figure 21 training performance results: electricity consumption

The actual, simulated results from EnergyPlus and predicted (by neural network mode) total energy consumption values for each simulation plotted in Figure 21 are almost a perfect match, with a loss value of 0.151 for normalized output values.

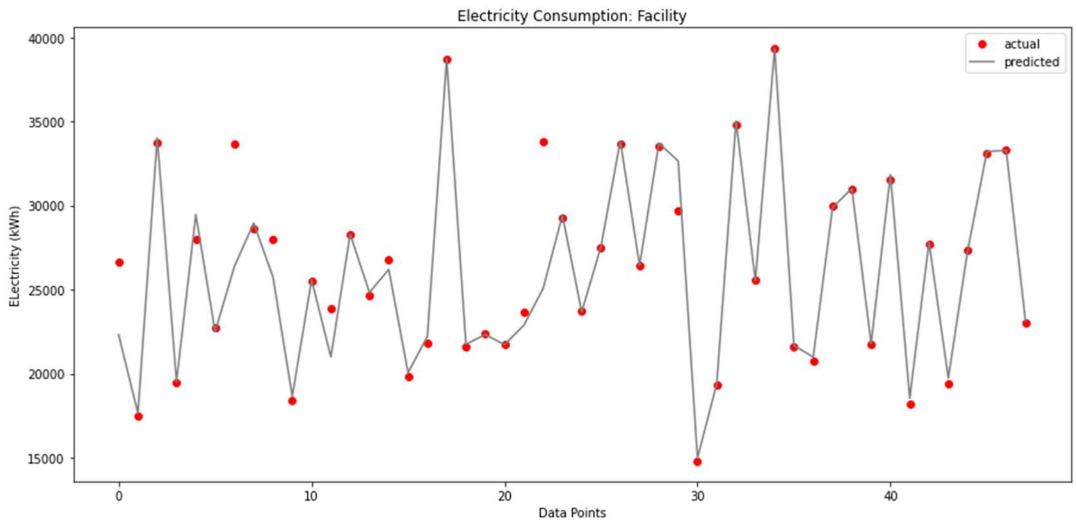

Figure 22 validation performance results: energy consumption

The model is then evaluated on the validation set. The model still performs well with a loss value of 1.4, and as expected, the loss would be higher than that of the training set. The lower loss implies a well-generalized model that can perform accurately on unseen data. Figure 22 proves that the NN model created can instantly predict building energy consumption, energy production, and net site energy for 50 different designs with great accuracy. The model still has outlying data points that are not accurately predicted. Some reasons causing the issue include sudden changes in building design energy consumption, and lack of enough training data to generalize over a wide range of building designs. Comparatively, EnergyPlus takes 1 minute average simulation time for just one building design. Note that the energy consumption values are relatively close to the reference model, which predicts a 26.7MWh annual power consumption. The co-simulation python program produces building designs with higher or lower consumption than the baseline (reference) model; because the same design values are generally applied to building components that share the same variable.

A similar comparison plot for the NN model prediction of the power production by the PV system and the actual BEM data shows a good correlation, as illustrated in Figure 23 below.

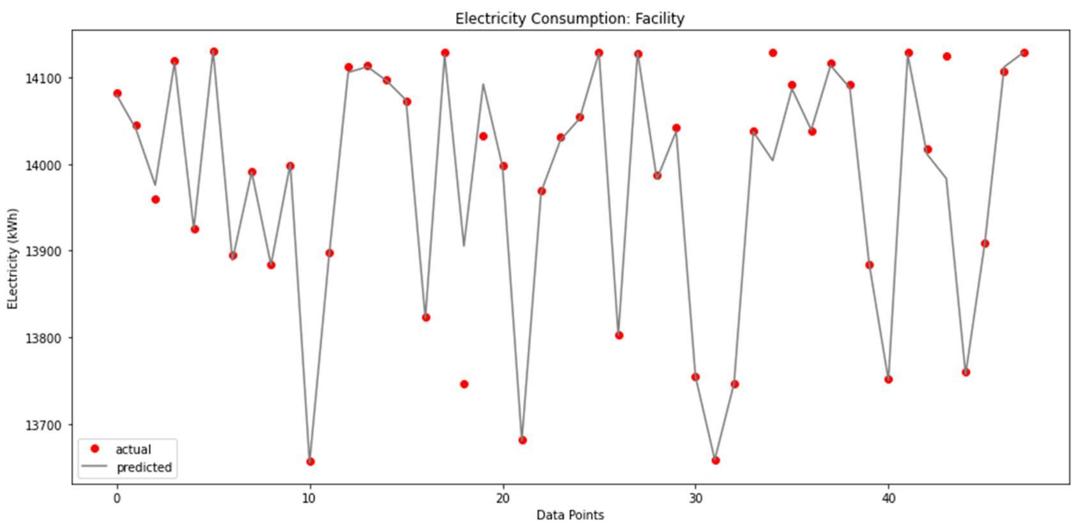

Figure 23 Actual and Predicted PV power production

The power production is only affected by one design variable: the PV panel tilt angle. The tilt angles are varied from 10 to 30 degrees, while the azimuth angle is 180 degrees. The results of power production move between 13MWh and 14.2MWh of electricity produced by the PV system in one year. The NN model has a relatively good prediction of the EnergyPlus BEM PV power production.

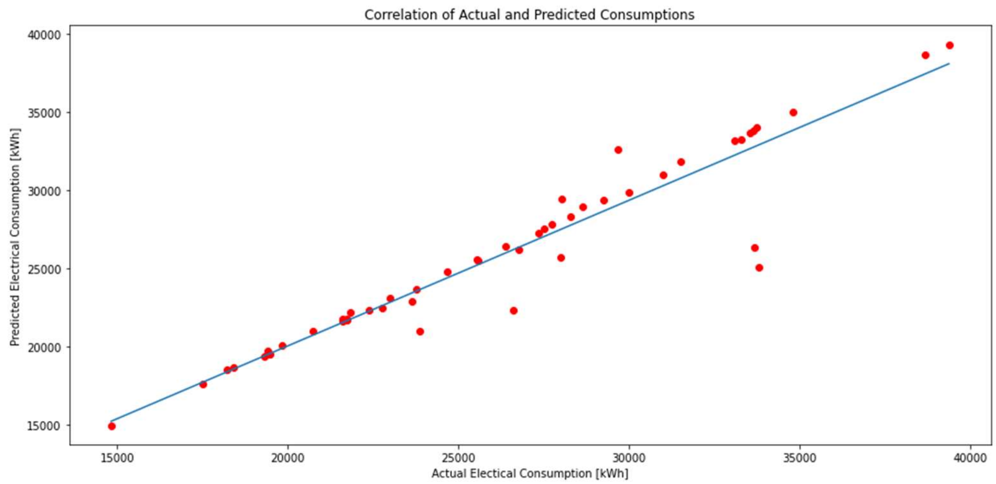

Figure 24 Correlation between EnergyPlus (true values) and NN-predicted energy consumption

An R^2 correlation coefficient was also used as a metric to measure how accurate the prediction model is doing on unknown data; the model achieved a maximum of 90% R^2 with different hyperparameter tunings.

v. K-fold Cross-Validation: Pandemic Surrogate Model

Better NN model training is achieved if there is no overfitting. Thus different techniques can be used to evaluate model performance and prevent overfitting. One technique used in this model is the K-fold cross-validation. Cross-validation is a resampling technique for evaluating NN models on a small sample of data. The process has one parameter, named k, which specifies the number of groups into which a given data sample should be divided. Thus the name k-fold cross-validation. When a precise value for k is specified, it may be substituted for K in the model's reference, for example, k=10 for 10-fold cross-validation. This study uses 5-fold cross-validation to determine model performance. Figure 25 is a convergence plot for the model when cross-validation is used on the dataset. Using K-fold cross-validation does not show a significant difference with the initial model. The minimum loss values for training and validation sets remain close to the initial model at 0.108 RMSE for training and 1.269 for the validation set.

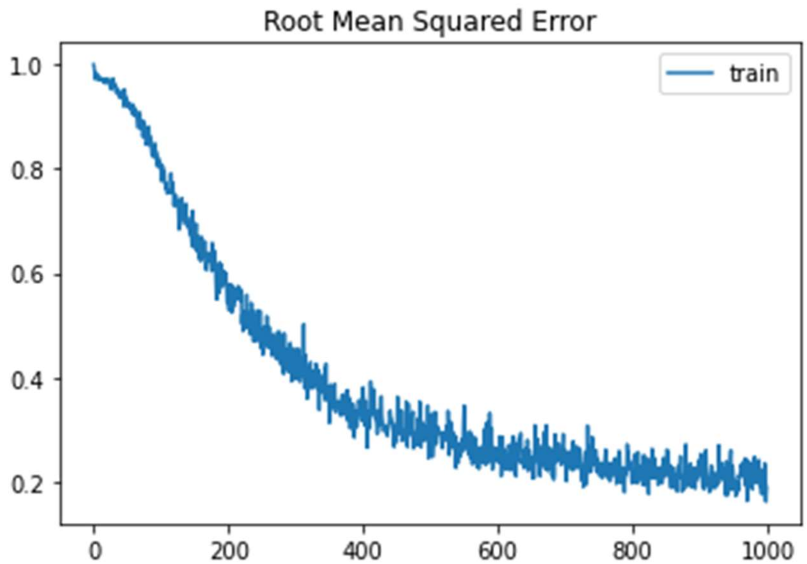

Figure 25 5-Fold cross-validation Model convergence

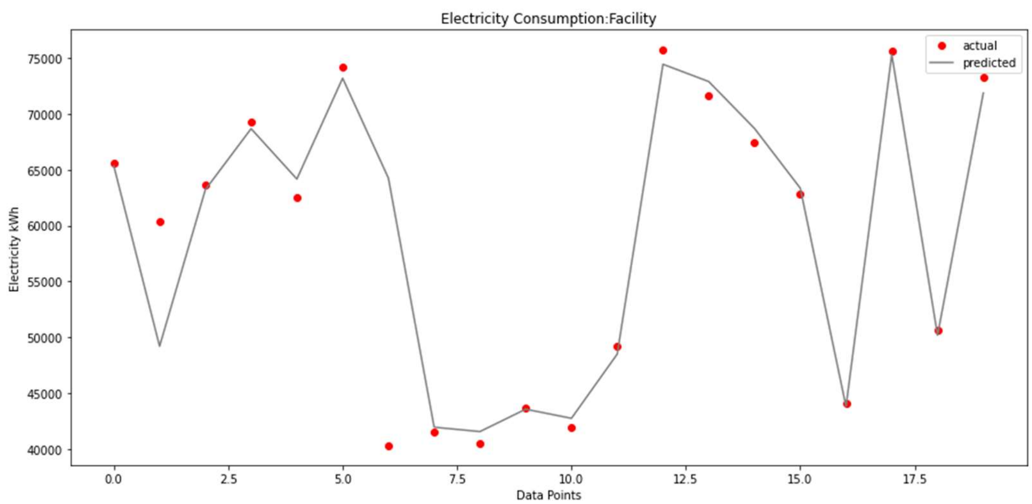

Figure 26 predicted and actual power consumption: K-fold validation

Results for actual and predicted building power consumption for 20 designs are plotted together for 5-fold cross-validation in Figure 26. These results are not different from those presented in Figure 22, with no cross-validation, confirming that the model is well-generalized.

vi. Surrogate Model for Regular Building Operation

Similarly, a model copy was used to train and test datasets developed for regular building operations. A good performance was achieved when the model was evaluated on unseen data representing this study's regular operation BEM model. This second model's hyperparameters were slightly modified to adapt to different data. Note, the goal was to make an NN-based surrogate model that can accurately predict the energy consumption of the FALC under regular operations. Suppose the model accurately predicts the energy consumption of this BEM. In that case, it is ready to be used as a surrogate model that can model FALC energy consumption, production, and net site energy. The surrogate model in this study can only predict building energy consumption based on 8 design variables whose values can be changed by the user. The surrogate model can be further improved to include more design variables. The EnergyPlus modeling scenarios make the surrogate model more complex but valuable for building energy designers in time and computational cost-effectiveness. Additionally, the model can simulate many different designs quickly, thus giving designers the option of making a fast parametric analysis to determine the best building energy design.

A dataset of two hundred data points of different building design models was generated from the co-simulation program. Figure 27, Figure 28, and Figure 29 are a representation of annual energy use, PV energy production, and net site energy across all generated design models. These values are slightly higher or lower than the reference model annual energy values. The reason is similar to the fact that the co-simulation program generally assigns the same values to all building objects sharing the same variables.

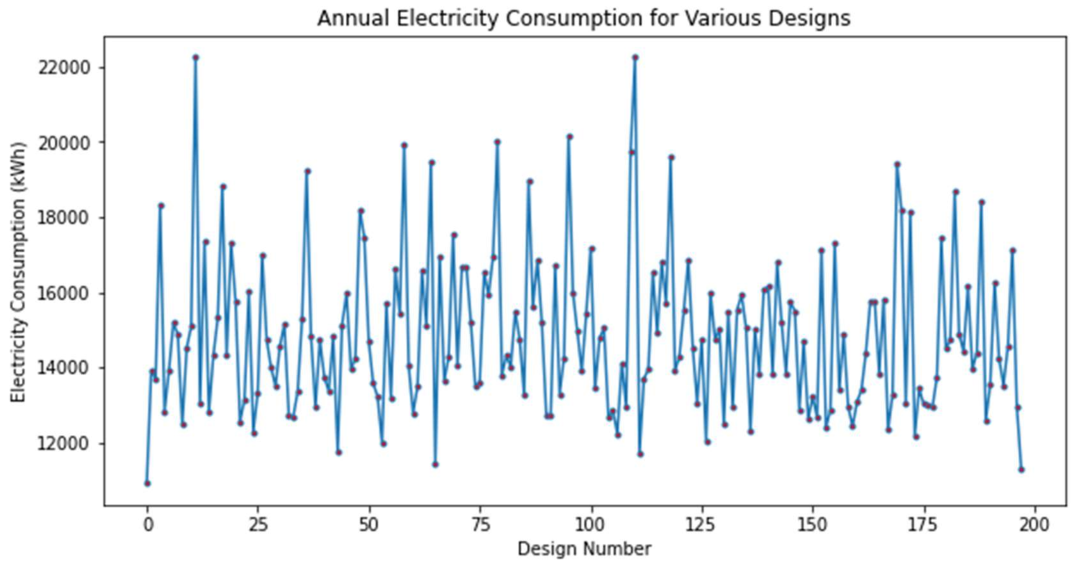

Figure 27 Electrical consumption distribution in the test set

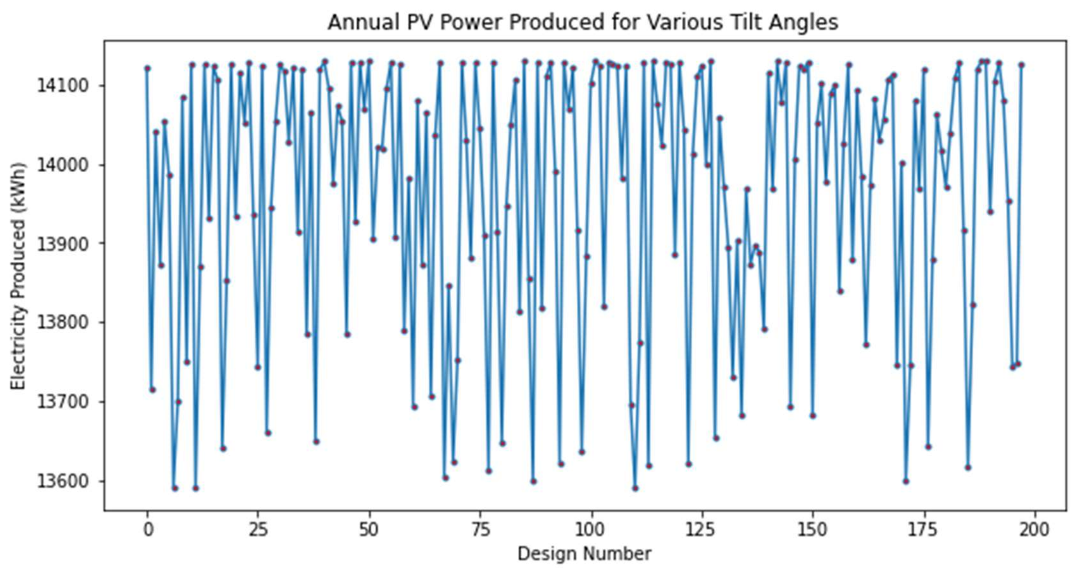

Figure 28 Test set data distribution for the PV power production

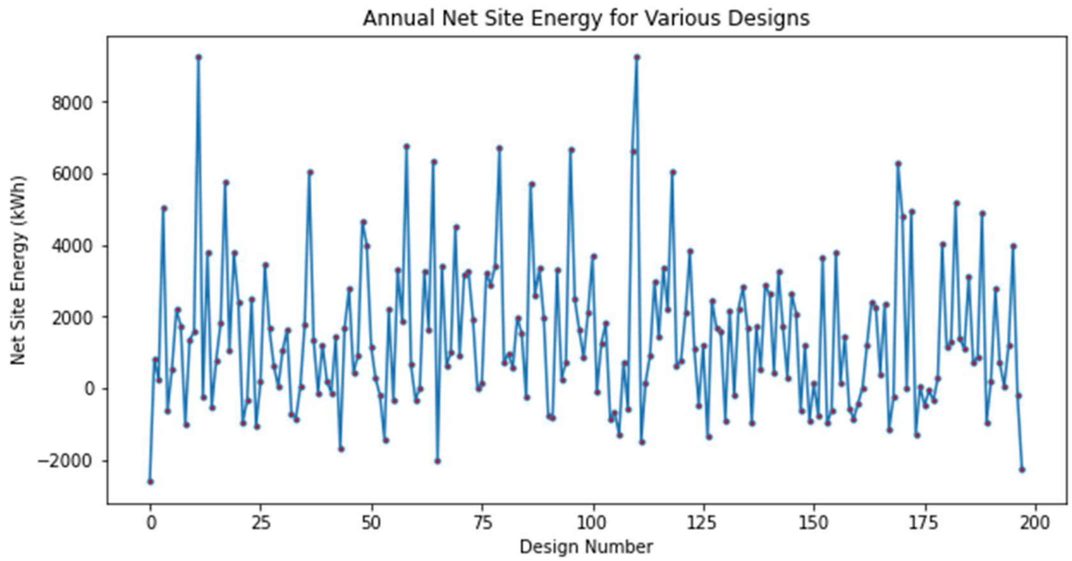

Figure 29 Test set data distribution for Net site energy

It should be noted that the python co-simulation program that uses EnergyPlus for simulations took 26 minutes in total to run all 200 simulations. That is slower than the few seconds taken by the NN model to predict energy consumption for the various designs.

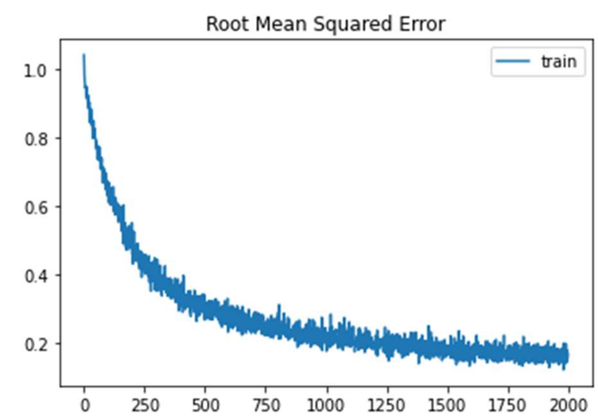

Figure 30 Model performance on the modified dataset

The regular operations model achieves an RMSE of 0.079 on training and 1.070 for the validation dataset. It should be noted that the learning rate of this model was relatively higher than the learning rate used in the pandemic operations surrogate model to achieve better performance. Figure 31 shows the comparison graph for predicted data and actual simulation data used for training. The results show that good learning is achieved.

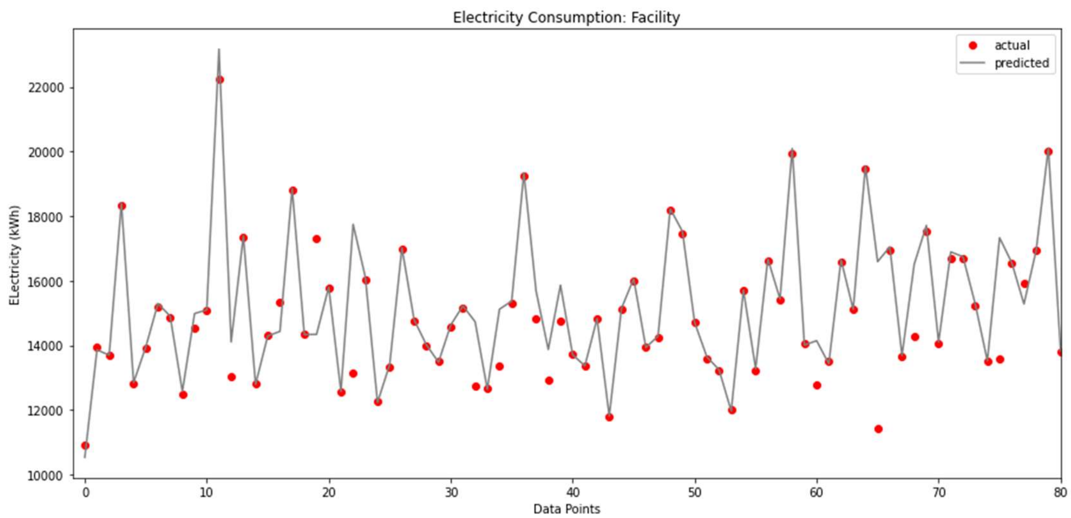

Figure 31 training performance results: electricity consumption

Similarly, a comparison graph is plotted in Figure 33, and the NN model accurately predicts almost all data points. The correlation graph plotted in Figure 33 also illustrates the similarity between NN model predicted data and simulated data from the test set. The majority of the data points show a perfect linear relation, indicated by the fit line, apart from a few data points away from the best fit line.

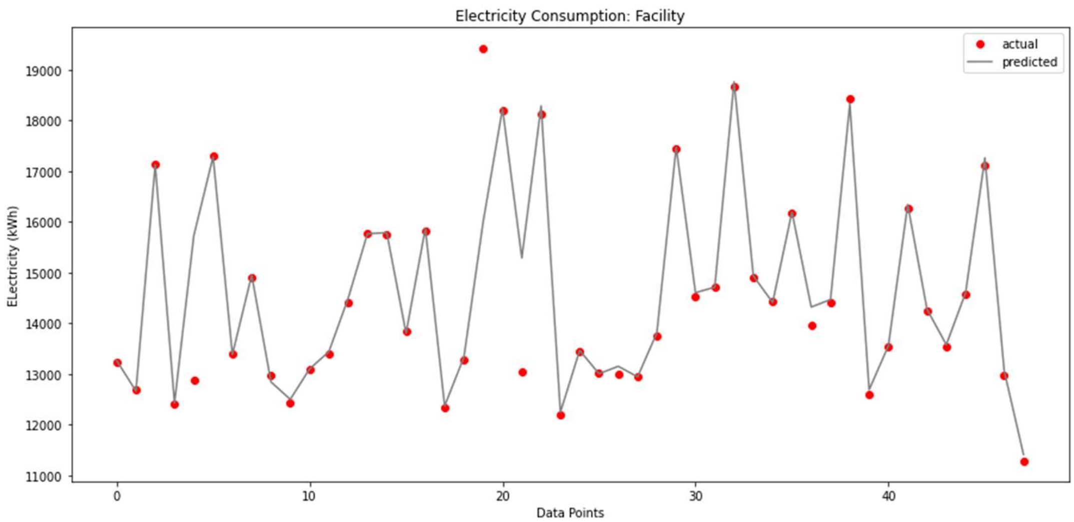

Figure 32 Actual and Predicted energy consumption

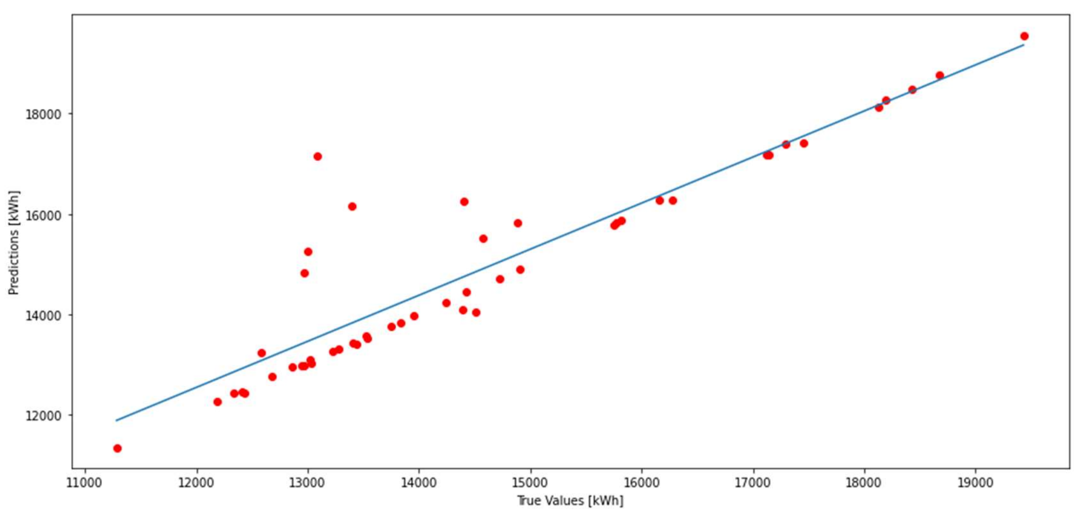

Figure 33 Correlation between EnergyPlus (true values) and NN-predicted energy consumption

The purpose of this test is to determine if this NN-based surrogate model is a good model that can be used as a function that represents the energy consumption, production, and net site energy. As a result, the surrogate model can be used for further studies and determine the best design variables that provide the minimum building energy consumption, as illustrated using genetic optimization in the next chapter of this study.

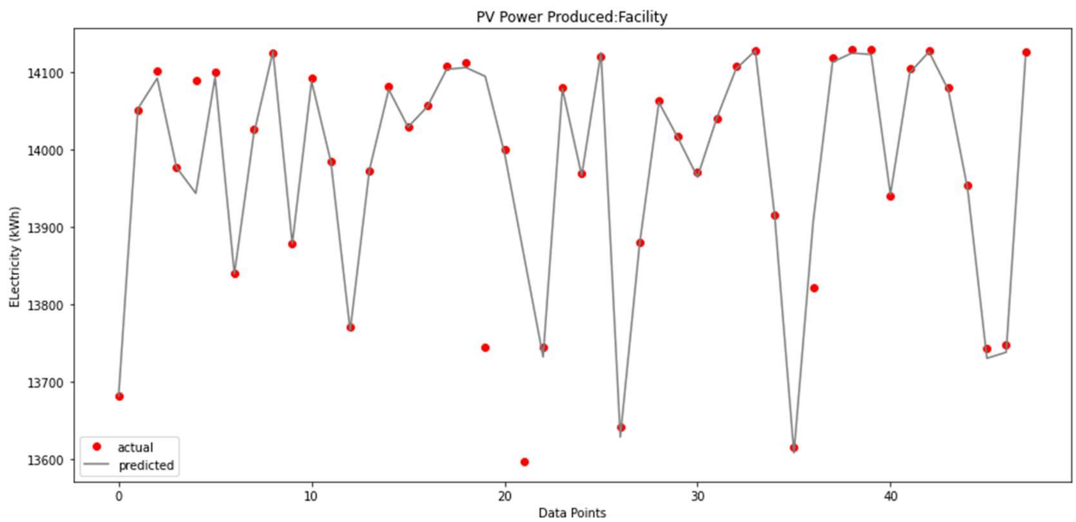

Figure 34 Actual and Predicted PV power production

The model also has a good performance in predicting the PV power production, as illustrated in Figure 34. However, cross-validation should be performed on the model for better evaluation given the few datasets being used.

vii. K-fold Cross-Validation: Regular Surrogate Model

5-fold cross-validation is used in this section as a way of validating the performance of this model. The K-fold cross-validation does not show a significant difference with the initial model. The minimum loss values for training and validation sets remain close to the initial model at 0.075 RMSE for training and 1.153 for the validation set. However, the predicted and actual data proved to have a better correlation, as shown in Figure 36 below.

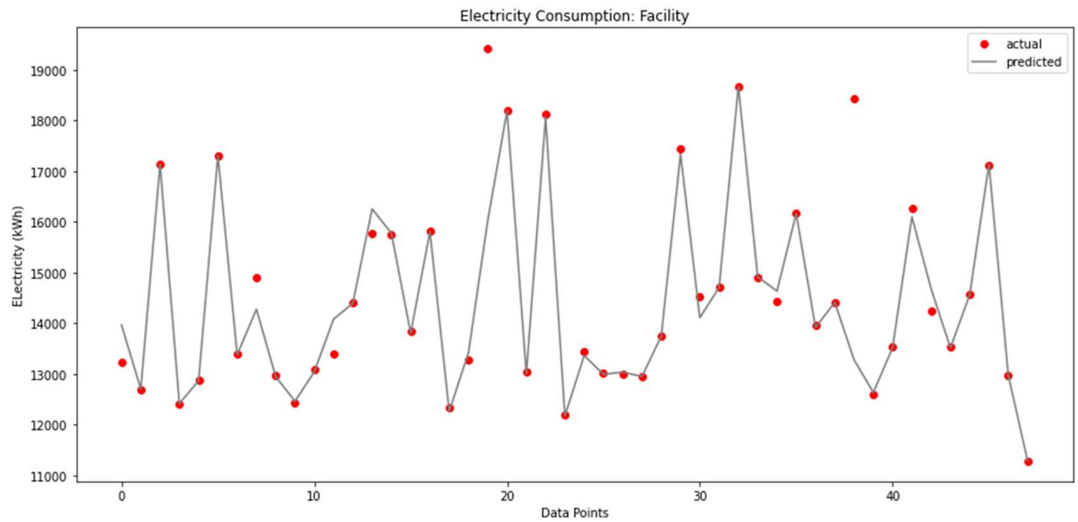

Figure 35 predicted and actual power consumption: K-fold validation

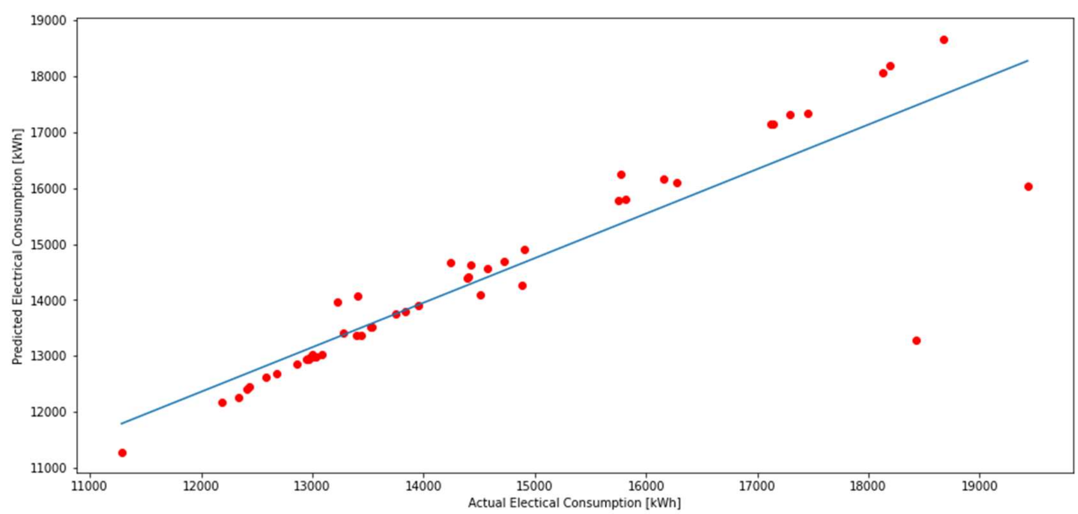

Figure 36 Correlation between EnergyPlus (true values) and NN-predicted energy consumption

Chapter 4
Optimization of the Neural Network-Based Model

i. Genetic Algorithm

The optimization described in this chapter is not the same as the optimization performed while developing the NN prediction model. The optimization is meant to find design variables that provide the minimum building energy consumption; this is done using a genetic function minimization algorithm. The NN surrogate model can only predict the energy consumption given a design variable inputs and doesn't modify the design variable. The surrogate model is enhanced when coupled with the GA to predict and select the best design variables with the lowest building energy consumption. Genetic algorithms make it possible to find a function's maximum (or a minimum), using operations analogous to those in the DNA of living beings: selection, crossing, and mutations [26]. Other methods can perform such an optimization, such as simulated annealing or conventional gradient descent methods. Still, the genetic algorithms have the advantage of resisting local minima reasonably well and have computational time advantages over simulated annealing.

Genetic optimization lies in the classification of algorithms known as evolutionary algorithms (EAs). Unlike traditional algorithms methods, the EAs are not static but dynamic, which means that they evolve. The evolutional optimization approach is appropriate for this study since the goal is to find the best optimal design that gives the lowest energy consumption with minimal computation cost.

Several energy efficiency-related studies have been made using a genetic algorithm to achieve building energy efficiency. A study that addresses bi-objective scheduling systems aiming to minimize electricity cost under time-of-use tariffs used a genetic algorithm in this study proved successful in achieving the study goals [27]. Similarly, warehouse energy management using a genetic algorithm is proposed to minimize energy consumption in warehouses in a study by Seval Ene et al. [28]. In their study, Ricardo Simões Santos et al.

used genetic algorithms to minimize energy consumption in buildings by determining the best choices of household appliances [29].

The Genetic optimization is implemented as shown in the flowchart presented in Figure 37:

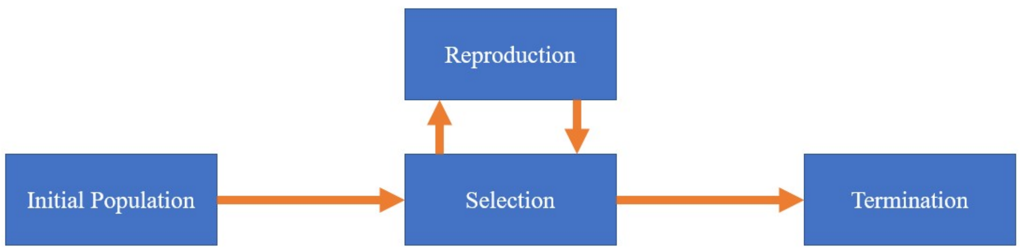

Figure 37 Genetic Algorithm Flowchart

The optimization process was carried out in 5 steps. The first step was population initialization, which is achieved by producing 100 different outputs from the surrogate model and saving the corresponding values for each design variable.

The second step is candidate selection. The selection process involves taking candidates from the population in the mating pool. Based on the previously calculated energy consumption value, a threshold (lower energy consumption than previously calculated) is used to select the best individuals. From the selected subset, parents are prepared for the mating. Selecting the parents to be used in mating is done using the Roulette wheel selection, which is an efficient way of selecting candidates by giving a higher probability of selection to the parents that have a lower energy consumption design model. The third step involves doing a crossover, where offspring are generated from two candidate parents. These offsprings have similar genetic properties as their parents, mathematically represented as decimal values in this study. Offsprings are then mutated to provide a different version with different genetic properties from its parent. New offspring are merged with the population, and the pool is sorted. Poor candidates are eliminated from the population. The remaining population is part of a new generation that undergoes a similar process starting from step 1 until a solution with the lowest energy consumption can not be overtaken. The process should be done in only 100 iterations; thus, either the minimum or energy consumption is reached, or the iterations are reached. The last step is terminating the process when the best solution is reached, and this can be compared as reaching the global minimum.

ii. Surrogate Model Energy Optimization

The minimization function approach is summarised with the equation below

$$\min f(x) = min\ of\ surrogate\ model$$

$$constained\ to$$

$$10 \leq x_1 \ = Wall\ R\ Value \leq 25$$

$$25 \leq x_2 \ = roof\ R\ Value \leq 45$$

$$2 \leq x_3 \ = HVAC\ COP \leq 5.5$$

$$1 \leq x_4 \ = Window\ U\ value \leq \ 7$$

$$0.1 \leq x_5 \ = Window\ SHGC \leq 1$$

$$21 \leq x_6 \ = Cool\ setpoint \leq 26$$

$$2 \leq x_7 \ = Light\ power\ density \leq 15$$

The GA optimization approach minimizes the energy consumption of the pandemic surrogate model. In other words, GA determines design variables that produce a building model with minimized energy use. The surrogate model is used as a fitness function to find a building design that may provide a lower energy consumption. These results can be compared to the results and design obtained when the minimization is done using the Bayesian Algorithm.

Comparatively, the Bayesian optimization uses a surrogate model as a black-box function to be minimized. An available package within python called Scikit Optimizer allows implementing the Bayesian optimization on the surrogate model. Thus this was worth implementing on the already developed surrogate model to compare the performance with the self-developed genetic algorithm; most of the inputs to this Bayesian algorithm are similar to those used in the genetic algorithm. Since the optimization variables considered in this study do not change for both the pandemic and regular operation models, the optimization illustration is only performed on the regular operation surrogate model, as shown in the next section.

iii. Regular Surrogate Model Energy Optimization

Once the NN surrogate model was modified and trained on the regular operations building design dataset, the model made good predictions that depict the actual BEM under regular operations. The surrogate model was then used as a fitness function for the genetic optimization and compared with those obtained with the Bayesian optimization. Constraints are applied during this optimization. Figure 38 presents the minimization trend followed by the GA algorithm for 100 iterations. The computation started converging after 40 epochs, and the total computation time was 5min 36s for a maximum of 100 iterations and an initial population size of 50 design models. Using more design models results in a longer computation time. However, since the computation should be as low as possible, the maximum iterations and number of design models are selected to provide the shortest time.

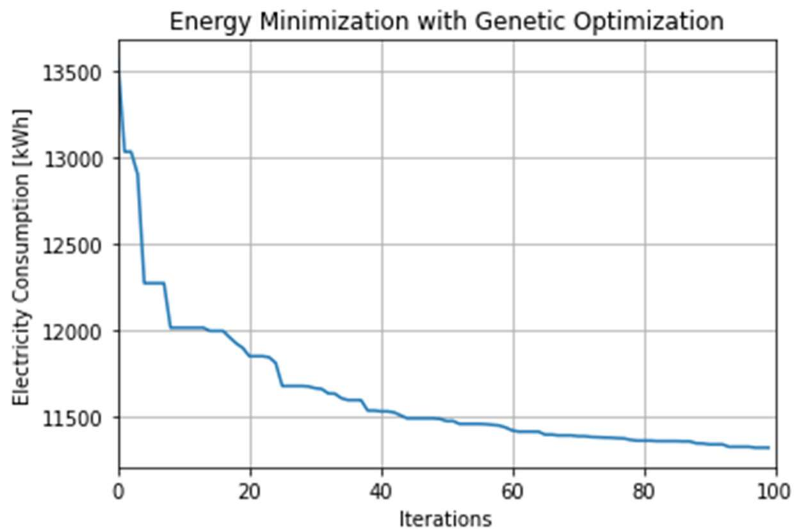

Figure 38 Building energy minimization with GA on test surrogate model

The GA minimizes the energy consumption of the regular BEM from 14MWh per year to 11.3MWh per year. A significant improvement in terms of energy consumption. The lower energy consumption is mainly achieved by the GA choosing a design model with more efficient design variables.

Table 16 GA optimization on actual BEM surrogate model

Parameter	***Best solution***	***Units***
Wall R-value	24.4	W/m^2K
Roof R-value	28.94	m^2K /W
HVAC Cooling COP	4.33	-
Window U value	4.75	W/m^2K
Window SHGC	0.40	-
Cool Setpoint	25.6	°C
Light power density	2.01	W/m^2K
Lowest Annual Energy Consumption	11,300.3	kWh

Design variables in this test model practically provide better energy efficiency. The cooling setpoint is in the range that would provide a lower energy consumption by the HVAC system.

The window U-value decision is considerably higher than the U-values for windows used in the building construction. This could result from the surrogate model errors resulting from a lack of proper window-U value and energy consumption correlation.

The recommended designs are then implemented in the EnergyPlus software to compare with the results from the surrogate model optimization. The EnergyPlus simulation results show a reduced energy consumption of 12420 kWh. Thus there is a 9.9% percent error between the optimization results using surrogate modeling and EnergyPlus modeling.

Likewise, the Bayesian optimization results are shown below in Figure 39, and tabulated results are presented in Table 17 for 100 iterations. The Bayesian optimization takes a shorter computational time compared with genetic optimization. The Bayesian optimization is a traditional optimization, unlike evolutionary optimization, and is expected to take a shorter computational time.

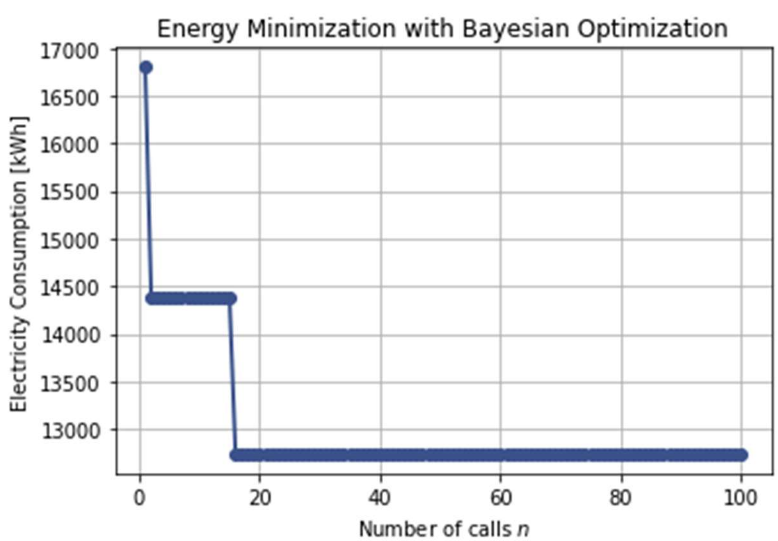

Figure 39 Bayesian Optimization on actual BEM surrogate model

The Bayesian optimization converges after 20 iterations and takes a shorter computational time. The total time for all 100 iterations was 59.5s; still, a faster computational time than that taken by the GA optimization. The Bayesian optimization can minimize building design energy consumption from 14MWh to 12.7MWh. A minimization that is not as low as the one achieved by the GA but does provide lower consumption than the actual BEM energy consumption estimates.

Table 17 Design variables obtained with Bayesian optimization

Parameter	***Best solution***	***Units***
Wall R-value	23	W/m^2K
Roof R-value	27	m^2K /W
HVAC Cooling COP	3.6	-
Window U value	4	W/m^2K
Window SHGC	0.44	-
Cool Setpoint	26	°C
Light power density	3	W/m^2K
Lowest Annual Energy Consumption	12,738.2	kWh

The Bayesian optimization results show an approach to lower energy consumption by increasing the wall R-value and the cool setpoint. However, the roof R-values and window U-values can be increased to have an even lower energy consumption.

The optimization process in this study creates a benchmark for a more complex optimization approach that involves the use of other variables such as variable electricity rates and systems life cycle costs etc. Additionally, the surrogate model could be improved in a way that provides the GA with multi-objective fit functions to optimize several variables at the same time. Such additional variables could be electricity cost, visual and thermal comfort.

The recommended designs by Bayesian optimization are then implemented in the EnergyPlus software to compare with the results from the surrogate model optimization. The EnergyPlus simulation results show a reduced energy consumption of 13,613 kWh. Thus there is a 6.86% percent error between the Bayesian optimization results on the surrogate model and EnergyPlus modeling.

Chapter 5
Conclusion

This study suggested a machine learning-based optimization strategy to predict and minimize overall energy consumption in buildings. Artificial Neural Network was used to create the surrogate model of an EnergyPlus based building energy model. This study combined surrogate modeling with one objective optimization to identify energy-efficient designs and their corresponding design variables. The approach was examined by reviewing the surrogate model error as a function of the simulated total energy value. There were noticeable prediction errors when the surrogate model was tested with an unknown dataset. The errors were specific to datapoints that have energy consumption values very high than the average energy consumption. The model was continually trained with more mixed datasets to improve the model's generalization.

A surrogate model proved to be more time effective in estimating building energy consumption compared to the traditional simulation software like EnergyPlus at the expense of an extensive training process to have a well generalized NN-based surrogate model. The surrogate model must be tested and modified multiple times before an accurate predictive model is achieved.

Energy optimization with genetic optimization algorithm is then performed using the surrogate model as a fitness function. The GA uses 8 variables with defined design space to perform function minimization that produces the least energy-consuming design model. Other methods can be used for optimizing the surrogate models. However, the genetic algorithm was chosen for its evolutionary approach to solving a given problem that tends to go a step further in finding the best solutions, unlike traditional optimization methods such as the gradients descent methods. Nonetheless, the GA optimization performance is compared with an easy-to-implement Bayesian optimization available in a python package to be easily integrated with the developed surrogate model. As expected, the GA optimization provided designs that can lower a building's energy consumption.

Similarly, Bayesian optimization also produces designs that lower a building's energy consumption. However, the GA takes a longer computational time than the Bayesian optimization; this is also expected as the GA takes an evolutionary approach involving many more steps to ensure the best solutions can be achieved. The GA optimization was able to provide lower energy-efficient solutions than the Bayesian optimization. An optimization performed on a surrogate model that presents the regular operations BEM for the FALC lowered the building energy consumption from 14MWh to 11MWh per year by choosing 8 design variables that provide more energy efficiency. Similarly, the energy consumption of a surrogate model representing the pandemic operations was lowered from 26MWh to 22MWh using the GA. The GA optimization made a balanced improvement to all design variables to obtain the best solution, unlike the Bayesian optimization that seemed to stick to those design variables that provide the best solution and keep making them better. On the other hand, the Bayesian optimization also lowered the regular operations surrogate model's energy consumption from 14MWh to 12 MWh per year.

In conclusion, the study successfully demonstrated the development and use of surrogate models to represent a building energy model. Furthermore, the study has successfully implemented building surrogate energy models and genetic optimization to find the best design solutions that offer an energy-efficient building design. Additionally, the whole process of generating an estimated building energy consumption, power production, and net site energy, given the required inputs to the surrogate model, takes less than 10 seconds. The surrogate model takes 3min 24s to predict 100 different design models based on the given inputs; this would be hours of simulation with traditional BEM simulation software. Thus, proving the time effectiveness of using surrogate models. Additionally, surrogate models can easily be connected to many optimization algorithms to determine the best design solutions, as illustrated with this study's GA optimization. The successfully created an approach for one to bridge between physics-based building energy models and strong optimization tools available in python which can allow achieving global optimization.

Future work should involve investigating techniques for sampling and training the surrogate model to decrease prediction errors and establish a clear correlation between all design variables and the target output to help the optimization algorithm find practical optimum designs. Additionally, the GA should be improved to allow multi-objective optimization.

References

[1] Chantrelle, F.P., Lahmidi, H., Keilholz, W., Mankibi, "Development of a multicriteria tool for optimizing the renovation of buildings.," *Applied Energy,* vol. 88, pp. 1386-1394, 2011.

[2] O. T. e. a. Karaguzel, "Coupling of Whole-Building Energy Simulation and Multi-Dimensional Numerical Optimization for Minimizing the Life Cycle Costs of Office Buildings.," *Building Simulation: An International Journal,* vol. vol. 7, no. EBSCOhost, doi:10.1007/s12273-013-0128-5., p. p. 111., 2014.

[3] M. e. a. Rabani, "achieving Zero-Energy Building Performance with Thermal and Visual Comfort Enhancement through Optimization of Fenestration, Envelope, Shading Device, and Energy Supply System.," *Sustainable Energy Technologies and Assessments,* vol. Vol. 44, no. EBSCOhost, doi:10.1016/j.seta.2021.101020, Apr. 2021.

[4] G. P. H. &. P. M.-O. Charles D. Corbin, "A model predictive control optimization environment for real-time commercial building application," *Journal of Building Performance Simulation,* no. DOI: 10.1080/19401493.2011.648343, pp. 159-174, 2013.

[5] W. Mcculloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biology,* vol. 52, no. (1-2), p. 99–115, 1990.

[6] F. B. M. a. H. D. O. Attneave, "The organization of behavior; a neuropsychological theory.," *Am. J. Psychol.,* vol. 633, p. 63, 1950.

[7] R. F. T. perceptron, "A probabilistic model for information storage and organization in the brain," *Psychol Rev,* Vols. 386-408, p. 65, 1958.

[8] B. W. a. M. E. Hoff, "Adaptive Switching Circuits," *RE WESCON Convention Record,* pp. 96-104, 1960.

[9] F. (. 1.). G. J. M. (. 1.). &. F. V. D. (. 1.). (. Bre, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks," *Energy and Buildings,* p. 1429–1441, 158.

[10] D. G. F. S. F. C. D. M. C. T. B. Guyot, "Overview of the use of artificial neural networks for energy-related applications in the building sector," *Int J Energy Res. ,* vol. 43, p. 6680– 6720, 2019.

[11] K. P. L. X. W. J. H. M. H. Li, "Multi-objective optimization for energy performance improvement of residential buildings: A comparative study," *Energies,* Vols. 10(2), p. 245, 2017.

[12] L. H. F. Magnier, "Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network.," *Building and Environment,* pp. 739-746, 2010.

[13] A. e. a. “. N. n. b. o. f. b. e. m. a. c. c. L. (. Jain, "NeurOpt: Neural network-based optimization for building energy management and climate control.," *L4DC,* 2020.

[14] Z. e. a. Jiang, "Building HVAC Control with Reinforcement Learning for Reduction of Energy Cost and Demand Charge," *Energy & Buildings,* vol. 239, no. doi:10.1016/j.enbuild.2021.110833., 2021.

[15] Y. e. a. Li, "An ANN-Based Optimization Approach of Building Energy Systems: Case Study of Swimming Pool.," *Journal of Cleaner Production,* vol. 277, no. doi:10.1016/j.jclepro.2020.124029., 2020.

[16] H. e. a. Li, "Building Energy Optimization Based on Biased ReLU Neural Network," *40th Chinese Control Conference,* no. doi:10.23919/CCC52363.2021.9549947., p. pp. 5933–5938, 2021.

[17] D. L. B. T. F. Gossard, "Multiobjective optimization of a building envelope for thermal performance using genetic algorithms and artificial neural network," *Energy and Buildings,* vol. 67, pp. 253-260, 2013.

[18] E. Asadi, M. G. d. Silva, C. H. Antunes and L. Dias, "Multi-Objective Optimization for Building Retrofit Strategies: A Model and an Application," *Energy and Buildings,* vol. 44, pp. 81-87., 2012.

[19] K. e. a. Bamdad, "Building Energy Optimization Using Surrogate Model and Active Sampling.," *JOURNAL OF BUILDING PERFORMANCE SIMULATION,* vol. 13, no. doi:10.1080/19401493.2020.1821094., p. 760–776, 2020.

[20] NREL, "USA_FL_Melbourne.Rgnl.AP.722040_TMY3," National Solar Radiation Database, Melbourne, 1991-2005.

[21] J. a. E. R. Gantz, " Extracting Value from Chaos," *IDC's Digital Universe Study,* pp. 1-12, 2011.

[22] B. P. Espinasse B, "Introduction au Big Data Opportunités, stockage et analyse des mégadonnées," Techniques de l'Ingénieur, Saint-Denis Cedex , 2017.

[23] "Training Neural Networks is ER-complete," *ArXiv abs,* vol. 2102.09798, p. n. pag, 2021).

[24] Y. B. J. L. P. L. Hugo Larochelle, "Exploring Strategies for Training Deep Neural Networks," *Journal of Machine Learning Research,* vol. 1, pp. 1-40, 2009.

[25] G. H. A. K. A. K. R. S. Nitish Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from," *Journal of Machine Learning Research,* vol. 15, 2014.

[26] M. H. A. a. S. K. Hamdy, "A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010," *Elsevier Science: ENERGY AND BUILDINGS,* vol. 56, p. p. 189, 2013.

[27] S. W. W. W. F. S. Kurniawan B, "Distributed-elite local search based on a genetic algorithm for bi-objective job-shop scheduling under time-of-use tariffs," *Evolutionary Intelligence,* Vols. v. 14, n. 4, p. 1581, 2021.

[28] K. İ. A. A. Ö. N. Ene S, "A genetic algorithm for minimizing energy consumption in warehouses," *Energy,* pp. 114:973-980, 2016.

[29] M. J. A. A. Santos Ricardo Simões, "Getting efficient choices in buildings by using Genetic Algorithms: Assessment & validation," *Open Engineering,* vol. Volume 9 Issue 1, pp. 229-245, 2019.

Appendix

Table 18:Pandemic operation dataset

Design #	Wall R value	Roof R value	HVAC COP	Window U value	Window SGHC	Cool setpoint	Light Pwr Density	PV Tilt angle	PV Energy	Building Energy Use	Net Site Energy
1	5	25	2	7	1	21	15	10	14122	10916	-2604
2	14	36	4	5	0	25	4	21	13715	13930	803
3	16	40	4	3	0	24	11	30	14041	13689	247
4	13	40	4	4	0	24	4	21	13872	18341	5062
5	20	32	4	6	0	22	8	27	14054	12829	-625
6	24	32	2	1	0	22	4	29	13985	13926	538
7	22	35	3	3	1	24	4	10	13590	15202	2195
8	12	41	3	5	1	22	11	27	13701	14862	1748
9	21	29	4	6	1	25	5	16	14083	12480	-1002
10	14	32	5	5	1	23	6	12	13749	14525	1365
11	18	40	4	5	0	24	13	21	14126	15090	1567
12	8	34	2	5	0	21	6	16	13589	22252	9246
13	17	28	5	5	0	23	6	21	13869	13042	-234
14	22	34	5	6	1	24	12	19	14125	17339	3817
15	14	32	3	6	0	22	11	10	13930	12804	-531
16	7	44	4	2	1	24	13	12	14123	14315	795
17	7	34	5	3	0	23	11	23	14105	15325	1821
18	25	29	4	2	1	21	10	13	13640	18812	5757
19	8	44	4	3	1	23	14	26	13852	14342	1081
20	12	28	5	7	0	22	6	15	14127	17323	3799
21	6	38	2	2	0	22	12	30	13933	15758	2421
22	8	32	5	2	0	22	10	17	14115	12555	-958
23	19	41	4	3	1	24	11	30	14052	13144	-309
24	13	34	2	4	0	21	14	24	14129	16043	2517
25	18	38	2	2	1	24	11	11	13936	12281	-1060
26	21	44	4	1	1	24	10	15	13743	13339	185
27	5	42	3	5	1	25	7	27	14123	16984	3464

28	19	26	2	6	0	23	4	17	13660	14739	1665
29	22	25	4	3	0	22	11	25	13945	14000	651
30	19	43	4	6	1	24	10	29	14053	13495	42
31	21	38	4	3	0	22	9	17	14126	14574	1051
32	7	26	5	3	0	22	4	13	14116	15160	1646
33	23	31	3	7	0	22	9	26	14028	12729	-700
34	19	40	4	3	1	24	11	11	14122	12684	-836
35	13	42	5	6	1	22	7	17	13914	13374	54
36	18	32	5	5	0	21	12	21	14120	15312	1795
37	21	27	4	5	1	23	8	26	13784	19236	6042
38	14	39	3	5	0	21	15	25	14064	14832	1368
39	8	44	5	6	0	22	10	20	13649	12939	-125
40	12	25	4	5	0	22	4	17	14119	14740	1224
41	8	41	4	2	1	22	6	25	14129	13737	210
42	18	44	2	3	1	23	11	14	14095	13371	-123
43	8	33	4	5	1	21	4	22	13975	14816	1437
44	8	44	4	3	0	22	4	11	14073	11778	-1694
45	6	33	3	1	0	25	14	25	14054	15127	1673
46	24	37	4	3	0	21	12	28	13785	15983	2788
47	6	39	4	3	0	23	7	23	14129	13953	427
48	17	40	3	3	0	22	3	18	13928	14252	919
49	16	31	5	2	0	24	7	22	14129	18201	4674
50	17	35	2	1	0	22	8	14	14070	17466	3997
51	15	26	5	5	1	22	7	28	14129	14702	1175
52	18	33	5	2	1	23	2	17	13906	13587	275
53	22	37	2	5	1	25	14	27	14020	13220	-201
54	15	33	2	6	0	23	9	22	14018	12008	-1411
55	8	31	4	3	1	24	13	28	14095	15696	2203
56	24	30	5	5	1	24	12	16	14128	13202	-323
57	20	29	5	6	0	23	10	20	13908	16620	3306
58	14	29	5	1	0	22	13	20	14126	15418	1895
59	15	32	3	6	0	25	13	23	13790	19950	6750

60	12	27	3	5	0	24	15	16	13982	14042	658
61	7	44	2	2	0	23	6	29	13694	12788	-319
62	12	42	2	7	1	22	4	14	14080	13496	17
63	11	35	5	6	1	23	3	19	13872	16566	3286
64	9	31	4	2	0	23	11	12	14065	15121	1657
65	23	44	5	3	1	22	12	22	13706	19472	6353
66	24	34	2	1	0	21	5	16	14037	11427	-2011
67	11	33	4	5	1	23	9	22	14128	16958	3433
68	23	42	2	3	1	22	10	12	13604	13641	620
69	22	28	5	2	0	25	5	20	13846	14276	1023
70	20	41	4	5	0	21	15	10	13623	17539	4500
71	11	34	4	2	1	23	15	15	13752	14063	900
72	7	26	3	5	1	22	3	11	14128	16684	3159
73	8	30	5	6	1	25	12	13	14029	16681	3251
74	15	43	2	5	0	21	3	27	13880	15218	1930
75	18	33	2	5	0	21	8	20	14128	13524	-1
76	8	32	4	6	1	22	6	16	14045	13575	129
77	24	30	4	2	0	23	3	27	13909	16545	3231
78	5	34	5	5	0	24	15	21	13613	15917	2888
79	18	30	3	1	1	23	14	16	14129	16942	3416
80	12	33	2	4	0	25	13	28	13914	20021	6702
81	15	25	2	6	1	21	3	17	13647	13792	730
82	6	26	5	3	1	22	13	11	13946	14321	971
83	22	30	5	6	1	22	10	17	14049	14011	561
84	12	43	4	5	0	21	15	21	14106	15494	1990
85	7	27	4	2	1	22	9	24	13813	14754	1532
86	15	26	2	2	0	24	9	14	14129	13284	-242
87	18	41	4	2	1	25	14	28	13855	18987	5724
88	6	45	2	3	0	21	4	15	13599	15619	2603
89	18	29	3	7	0	22	10	10	14129	16870	3344
90	16	42	3	2	0	21	5	14	13817	15192	1966
91	25	27	5	5	0	22	4	24	14110	12737	-772

92	12	37	5	5	0	23	7	29	14129	12729	-797
93	5	34	2	5	0	23	11	19	13990	16703	3310
94	9	25	5	2	1	22	15	11	13622	13298	260
95	18	36	4	7	0	22	11	27	14129	14260	734
96	11	40	2	6	1	24	5	22	14069	20162	6694
97	5	26	2	3	0	24	6	26	14122	16005	2486
98	6	31	4	5	1	25	10	17	13917	14957	1635
99	12	37	5	7	1	25	8	11	13637	13900	848
100	25	45	6	1	0	25	2	30	13884	15418	2127
101	12	34	4	5	1	22	6	21	14101	17188	3689
102	20	37	5	1	1	22	7	13	14129	13440	-87
103	5	27	4	4	1	21	10	22	14125	14780	1259
104	17	30	5	4	0	21	8	23	13820	15051	1821
105	11	40	5	7	0	24	12	20	14128	12663	-862
106	10	43	5	4	0	22	15	11	14127	12861	-663
107	7	39	5	4	1	23	9	25	14124	12241	-1280
108	6	41	2	2	0	22	6	17	13981	14122	737
109	20	25	3	5	0	24	7	24	14124	12951	-570
110	22	41	4	2	0	23	12	28	13696	19740	6631
111	24	26	3	7	0	24	3	30	13589	22252	9246
112	12	43	3	4	0	23	4	14	13775	11713	-1472
113	11	41	5	7	0	24	12	27	14128	13700	174
114	15	36	5	7	0	21	7	27	13618	13957	923
115	25	30	4	2	0	22	14	30	14129	16522	2996
116	20	32	3	1	0	23	10	19	14075	14909	1434
117	17	26	5	1	0	23	12	30	14022	16802	3378
118	9	27	2	6	1	23	12	12	14128	15730	2204
119	24	43	4	1	0	25	3	13	14127	19593	6069
120	23	31	4	6	0	21	14	29	13885	13916	624
121	18	36	5	2	1	23	10	10	14128	14303	778
122	23	44	3	6	1	24	11	28	14043	15540	2096
123	21	25	4	3	1	22	7	22	13622	16873	3835

124	17	27	3	6	1	24	14	20	14012	14530	1116
125	7	39	4	6	1	23	8	29	14110	13052	-456
126	8	43	2	4	1	22	14	27	14124	14728	1206
127	17	30	4	2	0	24	9	16	13999	12048	-1354
128	8	34	5	4	1	21	9	29	14129	15982	2455
129	8	37	2	5	0	24	6	11	13653	14748	1680
130	5	43	5	2	1	24	9	20	14059	15034	1575
131	24	32	5	6	0	21	10	24	13971	12487	-888
132	6	34	4	4	1	21	9	26	13895	15487	2186
133	22	41	5	3	0	23	5	19	13731	12938	-205
134	13	32	3	6	0	21	9	28	13904	15541	2231
135	17	25	5	6	1	22	5	11	13682	15935	2839
136	12	29	4	7	1	22	13	21	13968	15057	1685
137	7	43	5	6	0	23	12	18	13873	12320	-960
138	20	29	4	2	1	22	8	16	13896	15020	1718
139	13	32	3	5	0	24	7	16	13889	13838	543
140	10	44	2	6	0	24	8	12	13791	16063	2862
141	17	37	3	4	0	22	15	18	14114	16146	2634
142	17	38	5	5	0	23	15	16	13968	13831	459
143	7	45	3	2	0	22	13	16	14129	16786	3260
144	20	35	4	3	0	22	8	16	14077	15217	1740
145	16	29	2	2	0	23	7	14	14129	13844	318
146	9	35	3	2	1	22	10	25	13694	15748	2641
147	7	35	5	1	1	24	7	18	14005	15498	2090
148	11	42	3	4	1	25	6	28	14124	12884	-637
149	10	31	5	2	1	25	14	27	14119	14699	1182
150	22	29	3	4	0	24	15	12	14129	12612	-915
151	19	33	4	7	1	22	11	19	13682	13224	129
152	6	43	5	3	0	22	2	30	14051	12680	-771
153	17	30	4	7	0	25	12	25	14101	17144	3644
154	13	27	4	1	0	23	13	27	13977	12407	-973
155	21	30	5	2	0	21	13	12	14089	12866	-623

156	16	41	5	4	0	24	13	21	14099	17293	3795
157	16	41	3	3	1	22	7	24	13840	13392	143
158	20	38	5	2	0	23	12	18	14026	14890	1462
159	11	34	2	4	0	22	12	24	14125	12968	-554
160	9	39	5	5	0	22	12	15	13878	12430	-855
161	23	43	4	3	1	25	4	20	14093	13086	-405
162	19	26	4	3	0	23	8	30	13984	13404	17
163	14	42	4	5	0	23	13	16	13771	14396	1215
164	18	33	5	6	0	23	15	23	13973	15770	2394
165	22	29	4	3	0	22	5	19	14082	15751	2270
166	14	31	5	6	1	22	13	13	14030	13833	402
167	11	43	4	6	1	24	8	18	14057	15821	2365
168	10	33	2	6	0	24	13	23	14107	12342	-1163
169	17	32	3	4	1	24	7	21	14112	13280	-230
170	14	39	5	4	0	24	13	24	13745	19431	6275
171	18	41	5	3	1	24	8	25	14001	18195	4792
172	11	41	2	7	1	21	12	13	13598	13030	15
173	19	42	2	6	0	22	6	19	13745	18126	4969
174	11	30	4	3	0	21	8	10	14080	12191	-1288
175	23	33	3	5	1	23	10	13	13969	13439	67
176	6	39	5	6	0	22	7	22	14120	13025	-492
177	22	31	4	3	0	25	4	18	13642	12998	-59
178	11	39	5	5	0	22	3	25	13879	12944	-343
179	14	34	5	5	0	22	9	16	14063	13749	286
180	9	26	3	4	0	22	10	22	14017	17450	4032
181	10	34	2	5	0	23	7	20	13970	14515	1141
182	13	32	4	4	0	22	14	18	14039	14723	1283
183	24	36	3	2	0	24	4	21	14108	18677	5171
184	6	25	3	7	1	24	7	24	14128	14902	1378
185	16	39	3	5	0	21	14	29	13916	14419	1098
186	17	35	5	3	1	23	3	17	13617	16162	3129
187	10	43	3	2	1	24	2	10	13821	13955	725

188	23	34	3	3	0	25	5	14	14119	14398	881
189	18	36	2	1	1	23	5	28	14129	18424	4898
190	7	40	5	4	0	24	9	28	14129	12582	-944
191	9	30	5	5	0	22	14	17	13940	13537	193
192	19	44	2	7	0	24	10	24	14104	16273	2770
193	21	44	4	4	0	21	4	29	14127	14242	718
194	21	35	3	2	0	24	5	22	14080	13527	47
195	6	28	4	3	1	24	7	18	13954	14576	1218
196	20	35	2	4	0	23	4	13	13743	17124	3970
197	10	40	5	4	0	25	14	13	13748	12965	-195
198	13	28	5	1	0	25	7	29	14126	11285	-2237

Table 19 Regular operation dataset

Design #	Wall R value	Roof R value	HVAC COP	Window U value	Window SGHC	Cool setpoint	Light Pwr Density	PV Tilt angle	PV Pwr	Building Energy Use	Net Site Energy
1	5	25	2	7	1	21	15	10	14122	13584	57
2	14	33	3	6	0	25	9	20	13774	20382	7189
3	10	36	4	4	0	22	11	21	13843	24359	11099
4	5	28	3	3	1	24	8	13	14049	25360	11902
5	14	25	3	4	0	23	11	21	14079	26168	12682
6	13	41	4	4	1	24	6	19	13647	21582	8511
7	6	26	3	3	1	21	6	24	14129	23870	10336
8	5	33	5	2	0	24	2	13	13625	21768	8718
9	24	37	3	6	0	24	10	27	13677	17228	4128
10	11	30	4	3	1	22	4	13	14079	31034	17548
11	21	34	5	4	0	23	11	15	14017	27387	13959
12	16	34	4	1	1	25	10	21	13589	43462	30447
13	9	42	5	4	1	23	10	22	14072	21179	7698
14	22	43	3	6	0	22	3	11	13944	20495	7138
15	9	37	3	4	1	23	5	28	13710	22707	9576
16	22	32	2	6	0	23	2	11	13805	28857	15634
17	18	31	5	3	0	22	5	12	14076	20796	7312
18	10	34	4	6	0	21	14	22	14125	19700	6169
19	21	33	4	4	0	22	10	20	13621	23060	10014
20	13	38	3	2	0	24	6	22	14129	29862	16328
21	16	30	5	5	0	23	7	17	13848	21228	7963

22	7	35	3	6	0	21	5	12	14091	16551	3053
23	14	39	3	2	1	23	11	14	14023	27428	13995
24	12	39	3	4	0	21	4	22	13888	29628	16325
25	14	26	5	1	0	25	6	26	14035	34022	20578
26	18	26	5	2	0	25	10	11	14076	21797	8314
27	14	35	4	3	1	22	13	27	14125	28256	14726
28	14	32	4	5	1	24	4	15	14116	31168	17647
29	19	43	5	2	0	24	3	23	13630	19931	6877
30	18	37	2	6	1	24	5	16	13840	20321	7064
31	9	40	2	2	1	23	9	20	14119	21104	7578
32	25	38	3	4	1	22	3	22	14041	22513	9063
33	9	35	3	2	0	22	12	29	14096	31753	18251
34	14	44	3	6	1	24	10	25	14045	27804	14349
35	21	33	3	2	0	23	5	11	13831	17044	3796
36	23	25	3	2	0	24	4	15	13647	29409	16339
37	21	28	3	6	0	23	3	25	14059	26739	13271
38	24	38	3	6	0	24	5	21	14070	23749	10271
39	17	38	3	2	1	23	12	23	14028	17925	4487
40	23	37	4	4	0	23	2	15	13933	30422	17075
41	7	43	2	5	0	21	8	11	13965	25877	12500
42	6	33	4	5	0	22	10	21	13921	30467	17132
43	25	36	2	2	1	22	3	22	14034	20318	6875
44	16	44	3	2	0	22	3	20	13973	22463	9078
45	23	28	4	6	1	23	13	17	13780	26869	13671
46	21	34	5	3	1	25	12	18	13915	30056	16727

47	6	34	4	1	1	23	13	17	13631	22692	9636
48	6	40	5	3	0	23	6	20	13937	20849	7499
49	18	30	5	3	1	23	8	18	14129	25338	11804
50	8	30	3	4	0	23	12	17	14049	32214	18756
51	14	31	5	2	0	23	10	11	14129	25988	12453
52	21	26	3	5	0	24	2	17	14056	29293	15829
53	8	37	2	6	0	23	5	29	14094	35574	22074
54	7	40	3	2	1	25	12	21	14107	27946	14433
55	10	27	3	4	1	24	6	28	14064	27731	14259
56	8	30	4	5	0	24	14	21	14059	29126	15658
57	19	42	3	4	0	24	14	23	14123	15542	2015
58	12	35	5	5	0	21	14	24	13974	28540	15155
59	21	41	3	5	1	22	7	22	14057	25090	11625
60	18	37	5	3	0	23	3	30	14066	25481	12006
61	15	28	4	4	0	25	12	18	14128	28568	15035
62	24	34	4	2	1	23	9	21	14129	26160	12626
63	12	28	5	6	0	22	11	22	14123	25985	12457
64	21	37	3	6	1	22	10	29	14100	28664	15158
65	24	29	2	6	0	24	5	28	14116	32268	18746
66	8	40	3	6	0	21	7	30	13817	16630	3395
67	19	35	5	6	0	21	14	24	14002	22881	9468
68	11	27	3	5	1	23	12	25	14114	30731	17212
69	10	36	4	6	0	21	2	14	13892	19769	6462
70	7	39	3	6	0	21	14	25	13605	33401	20371
71	6	29	3	2	0	23	4	16	13893	31338	18031

72	25	29	2	4	1	24	11	10	14124	21645	8115
73	12	32	3	6	0	23	13	16	13630	27319	14264
74	21	40	3	2	0	24	6	26	14047	24250	10794
75	22	30	5	2	1	25	12	11	13829	20825	7579
76	14	29	5	6	0	23	10	21	14114	32747	19227
77	24	44	2	5	0	21	3	15	13981	25532	12139
78	18	40	3	6	0	23	13	25	14107	27072	13559
79	8	29	3	5	1	21	6	19	13897	20376	7065
80	24	29	4	1	1	23	4	16	14115	31849	18328
81	19	26	2	6	0	22	11	25	14120	18788	5263
82	22	31	4	6	1	22	3	25	14125	24835	11305
83	19	35	4	2	0	21	9	26	14082	23336	9846
84	19	31	3	4	1	23	5	23	14129	25053	11518
85	10	38	3	2	0	24	8	28	14097	31715	18212
86	22	32	4	7	1	22	13	23	13885	32144	18843
87	14	26	2	3	0	24	10	16	14128	20995	7461
88	25	29	5	2	0	23	8	27	13966	22054	8676
89	24	41	4	6	1	25	5	18	13753	16366	3192
90	23	32	2	3	0	24	10	15	13849	30401	17135
91	16	33	4	4	1	24	14	27	14128	31229	17696
92	24	28	2	3	0	24	14	28	14129	34379	20844
93	15	32	4	6	1	24	11	23	14094	27264	13763
94	9	26	3	3	1	23	7	26	14122	25510	11983
95	21	30	5	3	1	24	11	28	14129	26723	13189
96	21	42	4	4	1	21	8	16	13887	25994	12693

97	7	32	2	2	0	23	10	23	14086	28993	15499
98	9	37	3	5	0	24	5	24	14107	21935	8421
99	9	32	4	2	1	24	3	20	14017	17805	4378
100	25	45	6	1	0	25	2	30	14128	29152	15619
101	10	29	5	5	1	24	6	12	14129	21911	8377
102	11	30	3	7	1	22	12	11	14125	26121	12590
103	9	33	3	5	0	25	13	13	14035	27832	14388
104	15	27	3	7	0	21	10	17	13810	22542	9314
105	11	25	4	2	0	22	6	22	13918	16499	3167
106	16	42	2	3	1	23	11	14	13915	19426	6097
107	5	41	2	6	1	24	7	25	13621	28415	15370
108	6	33	5	1	1	24	5	29	14105	18729	5217
109	13	40	5	4	0	23	8	27	13714	22561	9426
110	20	37	3	1	1	24	9	26	14127	26640	13108
111	16	39	2	5	0	21	8	20	13589	43462	30447
112	13	28	4	2	1	21	7	14	14104	23993	10483
113	22	32	5	2	1	23	2	17	13710	25016	11885
114	13	36	5	6	1	24	3	17	13938	27451	14100
115	10	36	3	2	1	25	9	11	13776	24685	11490
116	10	31	4	6	0	25	3	24	13912	25733	12407
117	9	29	5	3	0	22	9	12	13885	25435	12135
118	17	33	3	2	1	23	7	29	14107	34688	21175
119	17	43	4	4	1	25	8	24	14122	17273	3746
120	25	40	5	2	0	24	12	12	14099	30623	17117
121	20	39	4	4	0	24	11	17	13689	23009	9897

122	13	40	4	5	0	21	9	13	13703	21137	8012
123	14	37	5	3	1	24	12	17	13595	24091	11071
124	8	28	5	4	0	24	11	16	14092	30620	17121
125	7	33	3	2	1	25	15	24	13787	27701	14496
126	19	27	5	6	0	22	3	26	14112	21860	8341
127	7	45	5	7	1	25	14	24	14114	18174	4654
128	18	36	5	6	0	24	10	12	13893	24938	11630
129	11	39	4	4	1	22	7	10	13874	26223	12934
130	11	43	3	2	0	22	11	23	13910	19814	6489
131	13	43	5	2	0	25	14	14	14060	26015	12548
132	20	37	2	5	0	22	2	25	14129	31263	17729
133	19	30	5	2	1	22	3	25	13655	31926	18848
134	17	31	4	1	1	24	9	16	13955	28619	15252
135	8	43	3	5	0	24	7	16	13628	18435	5382
136	10	29	4	4	0	24	5	17	13756	31140	17964
137	24	35	4	6	0	24	10	21	14008	17186	3767
138	11	36	4	4	1	23	15	27	13746	31636	18470
139	20	33	5	4	0	24	14	18	13633	21930	8872
140	19	44	3	2	0	24	2	11	14111	18969	5452
141	19	43	4	6	0	23	13	13	14129	28923	15389
142	14	39	5	4	0	22	4	19	14129	27516	13981
143	10	30	3	6	0	24	14	13	13818	35396	22160
144	16	38	3	6	1	22	4	11	13739	30884	17724
145	13	26	5	5	0	22	4	25	14051	31318	17859
146	8	43	3	4	1	25	8	28	13845	17095	3833

147	16	27	3	1	1	23	8	28	13777	20792	7596
148	23	31	2	6	0	21	13	14	14129	17718	4183
149	16	36	3	4	1	21	11	21	13756	19009	5833
150	9	30	5	4	0	24	3	15	14029	32591	19153
151	16	34	4	7	0	22	7	13	14082	26605	13116
152	24	33	4	1	0	25	5	28	14045	17495	4041
153	19	39	5	1	0	21	6	13	13960	33734	20362
154	25	33	3	4	1	22	12	20	14119	19489	5964
155	11	43	2	3	0	22	7	23	13926	28013	14673
156	18	25	5	6	0	21	4	21	14129	22755	9220
157	6	40	3	6	0	23	15	18	13894	33671	20362
158	12	32	5	5	0	22	5	25	13991	28629	15227
159	18	39	4	4	0	24	7	28	13885	28000	14700
160	25	35	3	3	1	24	15	16	13999	18395	4986
161	19	36	4	4	1	22	12	19	13657	25533	12452
162	19	29	5	5	0	21	15	16	13898	23845	10533
163	8	44	5	4	0	23	5	19	14112	28292	14775
164	22	28	5	2	1	24	11	11	14113	24651	11132
165	6	40	4	5	1	24	7	16	14097	26762	13258
166	6	41	5	1	1	24	14	25	14072	19822	6342
167	18	26	2	5	1	24	2	25	13824	21837	8596
168	24	35	4	3	0	22	12	24	14128	38667	25134
169	12	39	3	1	1	21	3	14	13746	21617	8451
170	8	43	2	4	1	24	14	29	14033	22364	8922
171	8	37	3	5	0	23	3	13	13998	21723	8314

172	6	42	3	3	1	22	3	20	13682	23625	10520
173	11	38	5	3	1	21	6	19	13969	33809	20428
174	8	45	4	4	1	24	6	12	14031	29233	15792
175	22	33	3	6	0	22	15	18	14054	23760	10297
176	15	31	3	4	1	24	9	20	14129	27504	13970
177	7	29	4	2	1	22	7	21	13804	33644	20422
178	6	44	4	6	1	22	9	29	14127	26388	12855
179	14	44	2	6	0	25	4	27	13987	33529	20131
180	7	25	2	4	0	21	11	19	14042	29678	16227
181	22	42	5	1	1	23	14	21	13754	14821	1647
182	17	39	5	3	0	25	2	13	13658	19304	6222
183	17	42	4	2	0	22	5	11	13748	34808	21641
184	23	35	2	6	0	25	10	13	14038	25581	12135
185	19	31	3	6	0	24	7	21	14129	39362	25828
186	17	34	2	5	1	23	14	28	14092	21598	8099
187	9	44	5	2	1	25	7	23	14039	20739	7291
188	15	43	5	2	0	21	8	21	14116	29978	16457
189	12	31	4	6	0	22	14	23	14092	30981	17482
190	18	26	4	5	0	23	7	16	13884	21718	8419
191	10	43	4	7	1	25	14	13	13753	31517	18344
192	13	28	5	6	1	23	3	29	14129	18204	4670
193	21	30	4	5	1	23	11	20	14018	27719	14291
194	13	25	4	1	0	23	4	30	14124	19398	5869
195	15	27	4	7	0	24	10	13	13760	27334	14154
196	9	42	2	1	0	25	12	16	13909	33087	19764

197	20	27	2	2	0	22	11	24	14107	33284	19771
198	8	26	4	7	0	23	7	29	14128	22985	9452

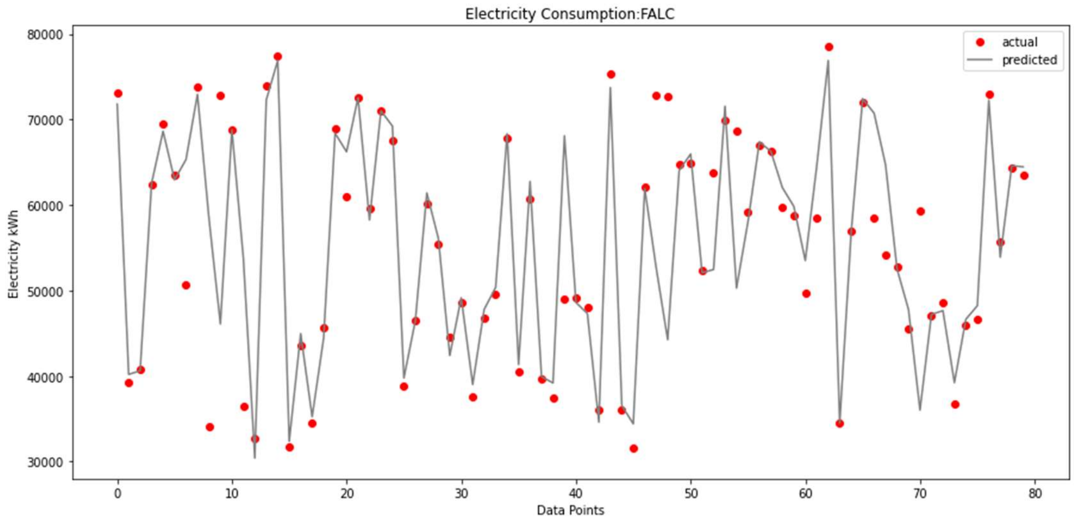

Figure 40 K-Fold validation on regular surrogate model: energy consumption

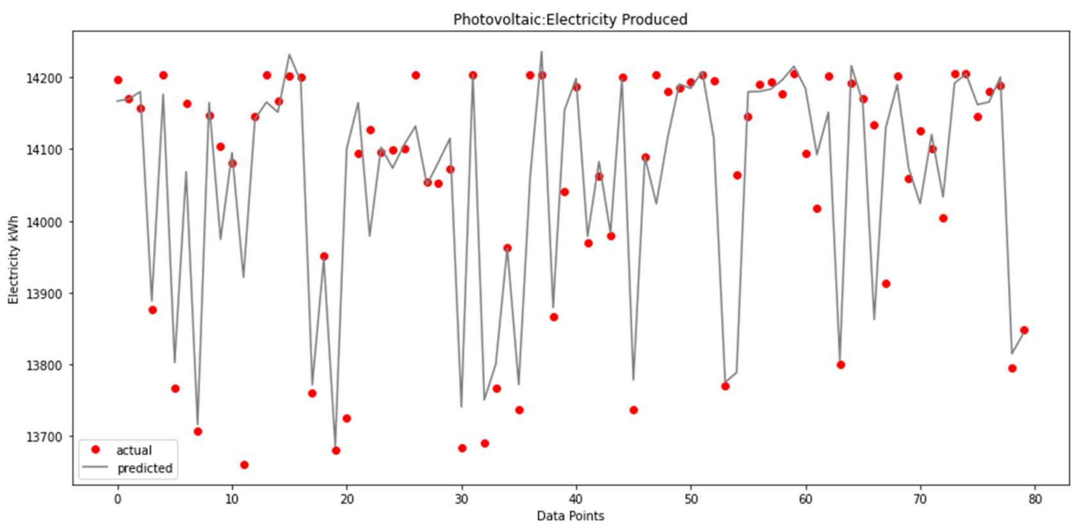

Figure 41 K-Fold validation on regular surrogate model: PV power production

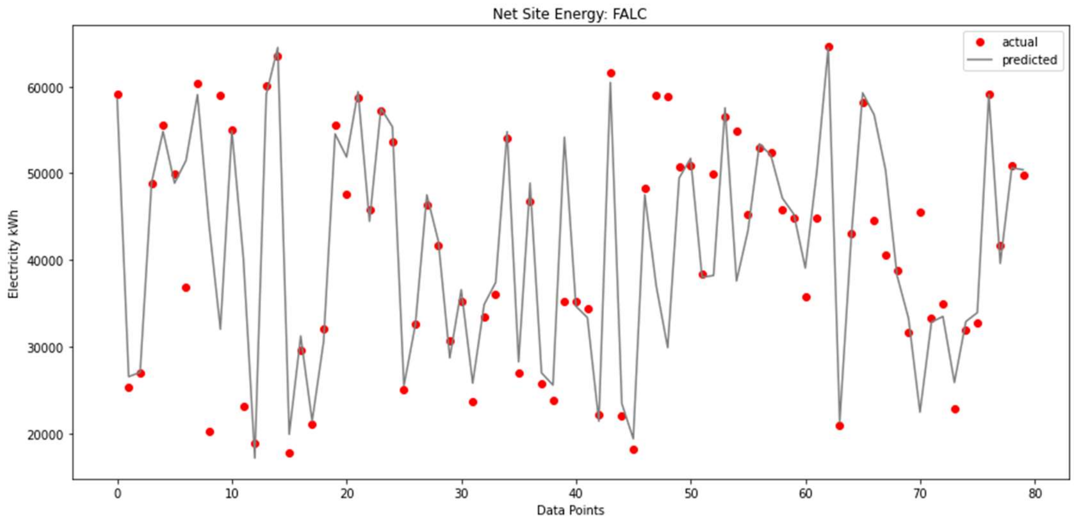

Figure 42 K-Fold validation on regular surrogate model: net site energy