buildings

Article

Artificial Neural Network for Predicting Building Energy Performance: A Surrogate Energy Retrofits Decision Support Framework

Haonan Zhang [1], Haibo Feng [2,*], Kasun Hewage [1] and Mehrdad Arashpour [3]

1 School of Engineering, University of British Columbia, Okanagan Campus, Kelowna, BC V1V1V7, Canada; haonan.zhang@ubc.ca (H.Z.); kasun.hewage@ubc.ca (K.H.)
2 Department of Mechanical and Construction Engineering, Northumbria University, Newcastle NE2 1XE, UK
3 Department of Civil Engineering, Clayton Campus, Monash University, Melbourne, VIC 3800, Australia; mehrdad.arashpour@monash.edu
* Correspondence: haibo.feng@northumbria.ac.uk

Abstract: Assessing the energy performance of existing residential buildings (ERB) has been identified as key to improving building energy efficiency and reducing associated greenhouse gas emissions in Canada. However, identifying optimal retrofit packages requires a significant amount of knowledge of building energy modelling, and it is a time-consuming and laborious process. This paper proposed a data-driven framework that combines machine learning, multi-objective optimization, and multi-criteria decision-making techniques to evaluate the energy performance of ERB and thereby formulate optimal retrofit plans. First, an artificial neural network (ANN) was developed to predict the energy performance of a wide range of retrofit packages. A genetic algorithm was employed to determine the best structure and hyperparameters of the ANN model. Then, the energy consumption results were integrated with environmental and economic impact data to evaluate the environmental and economic performance of retrofit packages and thereby identify Pareto optimal solutions. Finally, a multi-criteria decision-making method was used to select the best retrofit packages among the optimal solutions. The proposed framework was validated using data on a typical residential building in British Columbia, Canada. The results indicated that this framework could effectively predict building energy performance and help decision-makers to make an optimal decision when choosing retrofit packages.

Keywords: energy retrofits; artificial neural network; multi-objective optimization; TOPSIS

Citation: Zhang, H.; Feng, H.; Hewage, K.; Arashpour, M. Artificial Neural Network for Predicting Building Energy Performance: A Surrogate Energy Retrofits Decision Support Framework. *Buildings* **2022**, *12*, 829. https://doi.org/10.3390/buildings12060829

Academic Editor: Benedetto Nastasi

Received: 2 May 2022
Accepted: 8 June 2022
Published: 14 June 2022

1. Introduction

Energy use is the main factor influencing environmental sustainability and has received greater attention with the ongoing climate action initiatives. The building sector is responsible for over one-third of global energy consumption and nearly 40% of total direct and indirect greenhouse gas (GHG) emissions. The figures may double and even triple by 2050. According to the literature, building operation, control, and energy retrofits have garnered greater attention for their need to improve building energy efficiency and reduce associated GHG emissions [1,2].

In Canada, buildings are responsible for 12% of the total GHG emissions [3]. The government has launched many policies and standards to improve building energy efficiency in response to the increasing concern about building energy consumption and associated emissions. For example, the British Columbia Energy STEP Code (BCESC) was introduced to reduce 80% of energy consumption for new buildings by 2032 [4]. However, a lack of guidelines promoting the energy efficiency of existing buildings is the main challenge to reducing GHG emissions. Existing old buildings usually consume more energy and produce more GHG emissions because of deteriorating building materials and less-efficient building energy equipment [5]. As such, building energy retrofits have been identified as key to improving the energy performance of existing buildings and realizing climate mitigation goals. Furthermore, energy retrofits can deliver other benefits, such as reduced utility bills and enhanced thermal comfort.

Generally, building energy retrofit measures can be categorized into two clusters: demand-side measures and supply-side measures [6]. Demand-side (DS) measures refer to strategies to reduce heating and cooling load by enhancing the energy performance of building envelopes (e.g., improvement of wall and roof insulation) and energy equipment (e.g., replacement of the heat pump). The DS measures are considered the primary solutions to improve the energy efficiency of existing buildings [7]. On the other hand, supply-side (SS) measures usually include the installation of renewable energy equipment (e.g., solar photovoltaics) that are used to produce electricity for buildings [8]. In recent years, SS measures have been recognized as essential strategies to transfer conventional buildings into net-zero energy buildings [9].

The whole process of energy retrofits requires continuous efforts, including assessing possible energy conservation measures and identifying optimal retrofit solutions. Although many existing building energy modelling tools, such as TRNSYS (University of Wisconsin, Madison, WI, USA), EnergyPlus (Lawrence Berkeley National Laboratory, Berkeley, CA, USA), and HOT2000 (Natural Resources Canada, Ottawa, ON, Canada), can evaluate the energy-saving potential, they require much professional knowledge, high computational complexity and burden, and a large amount of energy simulation time to simulate the energy performance of retrofit scenarios [10]. This problem will intensify when there are a large number of retrofit scenarios to be simulated for more complex buildings. Given this issue, previous studies generally select a reference building to represent a group of buildings that have similar building characteristics (e.g., building envelopes and energy systems) and apply multi-objective optimization methods to identify optimal solutions. This can help decision-makers (DMs) avoid simulating many buildings and thereby save a significant amount of computational time [11]. For example, Pilechiha et al. [12] developed a simulation-based optimization framework for designing office windows. This framework can assist DMs in determining the best model that has the highest view and daylight and medium energy consumption. Similarly, Sharif and Hammad [13] performed multi-objective optimization for a reference institutional building to identify optimal retrofit solutions. The proposed framework can simultaneously minimize building energy consumption, life cycle costs, and environmental impacts. Similar research has been conducted by [14–16].

While the proposed approach is feasible considering computational burden and time, the results may not be robust for every single building of the investigated building group because a reference building may represent only a limited part of the group [17]. In reality, DMs should be aware of optimal retrofit scenarios for every single building or very similar buildings to take into account the building peculiarities and DMs' actual needs. To deal with this problem, researchers have proposed two approaches. The first one reduces the level of details of building energy models to improve work efficiency [18]. However, this can lead to inaccurate energy simulation results. The other generally develops a surrogate model to substitute traditional energy-simulation processes and thereby provide energy consumption results for retrofit decision-makers [19].

Armed with big data, powerful computing capacity, and advanced algorithms, machine learning (ML) has been increasingly employed in building energy research to help automate, simplify, and generalize the process of evaluating energy conservation measures [20]. ML algorithms have presented great advantages in processing a large quantity of data and making predictions based on existing data without background information [21,22]. Well-trained ML models can interpret the relationship between different variables by fitting a sophisticated function to a given dataset. Based on different training approaches, ML algorithms can be grouped into three categories: supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL). SL algorithms include a training procedure using labelled existing data to make a prediction, which can be used for regression or classification of unseen data [23]. UL algorithms can be employed to identify potential patterns in the unlabelled data for clustering the examples [24]. Different from SL and UL algorithms, RL control consists of an environment that approximates real-world operation and an agent that aims to achieve optimal control over the system. RL can find maximum rewards by identifying an optimal mapping from states to actions [25].

Among various ML algorithms, the artificial neural network (ANN) has been recognized as one of the most powerful surrogate models for evaluating upgrades in building envelopes and energy systems and determining optimal retrofit solutions [26]. For example, Thrampoulidis et al. [19] developed an ANN model for evaluating different retrofit measures at an urban level. The proposed model was tested through a case study in the City of Zurich, Switzerland. Results indicate that the proposed method can substantially reduce computational costs without compromising accuracy for most retrofit dimensions. Similarly, Beccali et al. [27] developed an ANN-based decision support tool to assess energy performance and refurbishment actions for non-residential buildings in Southern Italy. The results demonstrate that the proposed model can be used to predict the energy performance of buildings and the first selection of retrofit measures. Furthermore, Seyedzadeh et al. [28] proposed a data-driven model to provide a fast and accurate prediction of energy loads. The model can be used as an ideal tool for decision-making tasks related to building energy retrofit planning. Similar studies have been conducted by [29–31].

Although previous studies have developed several models for predicting energy performance and retrofit scenarios for buildings, few of them considered carbon emission and cost impacts when making retrofit decisions for Canadian residential buildings. In addition, Canada needs prediction models for the retrofit scenarios of its existing residential buildings, since over 50% of Canadian residential buildings are more than 30 years old, and many of them need energy retrofits to reduce carbon emissions [5]. Thus, this research proposed a comprehensive data-driven framework that can estimate the energy performance of a wide range of retrofit scenarios and help decision-makers select optimal retrofit packages to reduce carbon emissions at a relatively low cost. This article proceeds as follows: Section 2 describes the overall methodology of this research, Section 3 presents a case study in Canada to demonstrate the efficiency of the proposed framework, Section 4 shows the results of the case study, while Section 5 concludes the main findings of this research.

2. Methodology

In this study, an ANN-based energy retrofit evaluation framework was developed to evaluate the energy performance of a wide range of retrofit packages and identify optimal retrofit solutions. The developed framework mainly consists of three steps: (1) develop an ANN model as a surrogate model to imitate the building energy modelling process and thereby estimate the energy consumption of retrofit packages; (2) calculate carbon emissions and energy costs by integrating emission and cost impact data with energy consumption results. Then, utilize a multi-objective optimization approach to determine Pareto optimal retrofit solutions while considering the emission and cost performance; (3) apply the multi-criteria decision-making analysis to evaluate the optimal retrofit solutions and select the best retrofit packages. The overall methodology of this study is shown in Figure 1.

2.1. Artificial Neural Network

Empowered by big data, ML algorithms have been increasingly employed for prediction in various engineering-related fields [20]. This paper employed an ANN to predict building energy consumption and thus evaluate the environmental and economic performance of possible retrofit packages.

2.1.1. Data Collection and Preparation

Collecting data on existing buildings related to energy retrofit options is essential to developing an accurate ANN model to predict building energy consumption. The input variables include different retrofit options, such as upgrades in building envelope components and heating, ventilation, and air conditioning (HVAC) systems, while the output variables are natural gas and electricity consumption. Since the database on possible retrofit packages is still missing in Canada, this study collected data through a simulation-based approach.

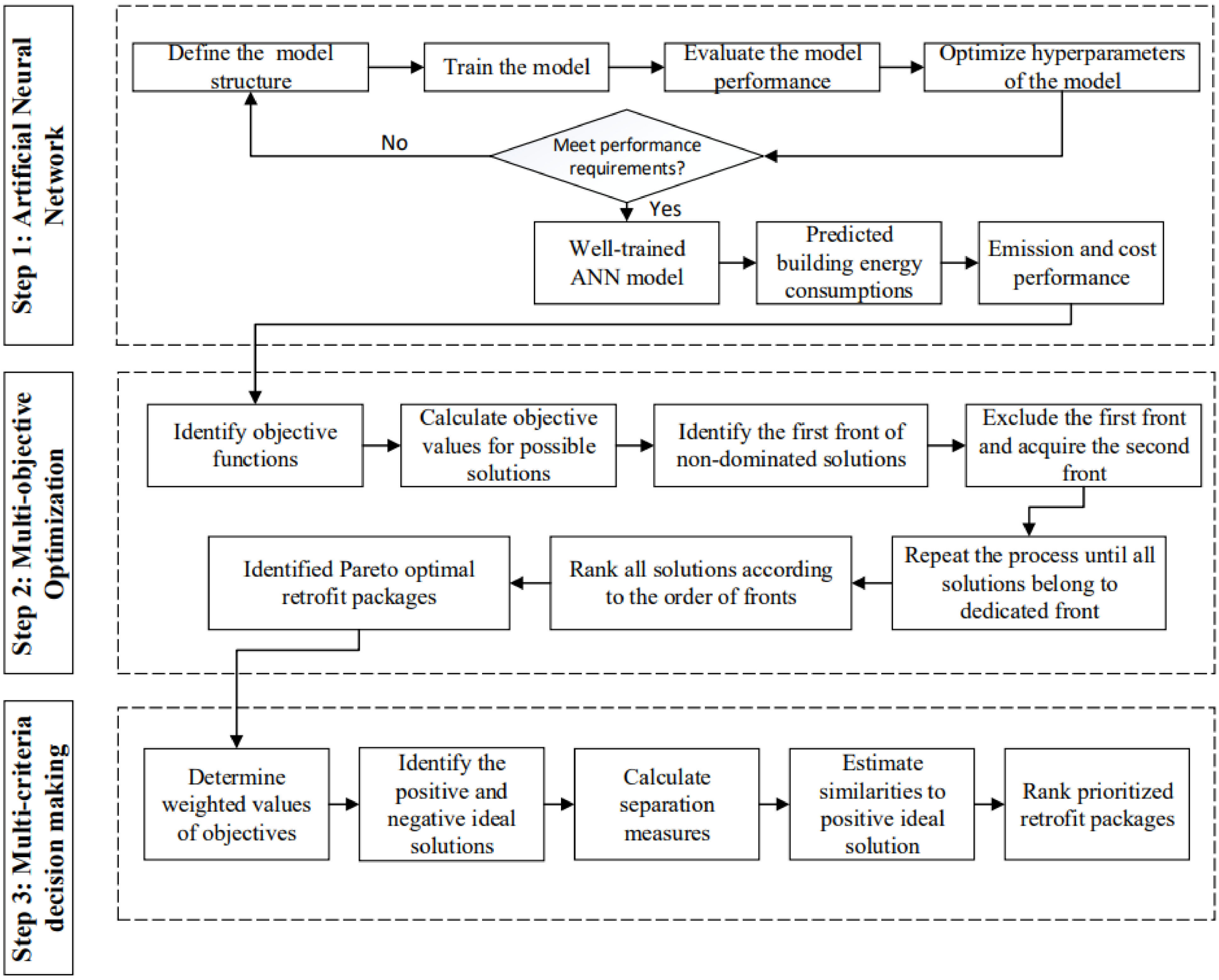


Figure 1. The overall methodology of the study.

In addition, data preparation is also important for developing an accurate ANN model. Preprocessing data can eliminate invalid values and inconsistencies for different variables through transformation and integration, thereby minimizing biased data and producing a complete dataset. In this study, the input variables can be grouped into two categories: quantitative variables (e.g., R values of building envelopes) and qualitative variables (e.g., heating equipment type). The former can be easily used to develop an ANN, while the latter needs transformation for the model development. This study employed the one-hot encoding approach to transfer qualitative data to quantitative data. In addition, in order to avoid the overflow error resulting from different magnitudes of input variables, the input values considered in this study are normalized using a linear transformation approach.

2.1.2. Model Development

An ANN is a parallel computational model enlightened by the human brain that comprises interconnected information processing units (named neurons) [32]. To deal with complicated tasks, an ANN performs a process of training where it adapts the characteristics of their interconnections to achieve a desired objective. There are different types of ANN models, such as the multi-layer perceptron (MLP) neural network, convolutional neural network, and recurrent neural network. Previous studies have proved that the MLP neural network is the most powerful and convenient ANN model for prediction, provided that there are multiple input variables and output variables [27,33,34]. An MLP neural network generally consists of one input layer, one or more hidden layers, and one output layer, as shown in Figure 2.

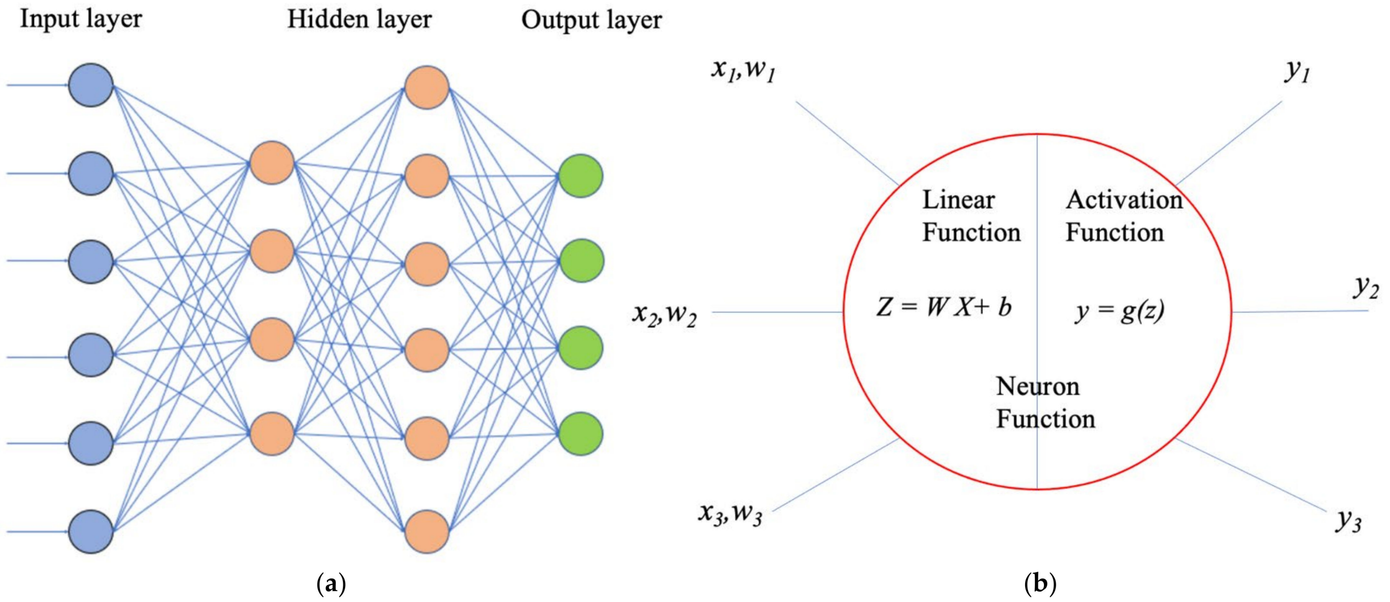


Figure 2. The structure of an artificial neural network: (**a**) the structure of neural network; (**b**) neuron function.

Each layer consists of an array of neurons, and their components are described as follows:

- Inputs x_i: The input features of the ANN model.
- Weights w_k: The parameter that needs to be identified by training the model.
- Bias b_k: The parameter that needs to be identified by training the model.
- Activation function g : The function of a neuron that defines the outputs of that neuron given a set of inputs.
- Outputs y_k: The output variables of the ANN model.

An ANN can learn the relationships between inputs and outputs by training historical data and calculating prospective outputs given input data [35]. Each neuron receives input data from neurons of the previous layer through weights and bias and then combines them by means of an activation function to produce output data. The development of an ANN includes the following steps: (1) data preprocessing: clean the provided data and spilt training and test data; (2) model design: identify the ANN structure (e.g., number of layers and neurons); (3) training: establish the relationship between inputs and outputs based on historical data; (4) evaluation: evaluate the model performance via a criterion (e.g., mean squared error).

The main objective of training an ANN model is to find optimal parameters (weights and biases) to minimize the cost function for the model [26]. However, it is difficult to identify the optimal values with multiple hidden layers [36]. Thus, a backpropagation algorithm was developed to deal with this problem. In addition, hyperparameters, such as the number of layers, neurons, activation functions, epochs, dropout rate, optimizer, and batch size, should also be identified when designing neural network structures.

2.1.3. Model Performance Evaluation

The goal of ANN is to build a model to predict an unknown case based on the training dataset. To assess the accuracy of a trained model, researchers need to compare predicted values with actual values on a test dataset [37,38]. Several evaluation metrics can indicate how accurate the developed model is. In this article, root mean squared error (RMSE), mean absolute error (MAE), relative squared error (RSE), and the coefficient of determination (R^2) were used to present the accuracy of the trained model. R^2 is the percentage variance in the dependent variable explained by the independent variables. These values can be determined by the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_i^n |y_i - \hat{y}_i| \quad (2)$$

$$\mathrm{RSE} = \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \overline{y})^2} \quad (3)$$

$$\mathrm{R}^2 = 1 - \mathrm{RSE} \quad (4)$$

Here, y_i, $\hat{y}_i$, and $\overline{y}$ are the actual, predicted, and average response values, respectively.

There are two main model evaluation approaches: (1) train and test on the same dataset; (2) train and test split. Training and testing on the same dataset can produce a high training accuracy. However, high training accuracy may lead to an over-fit of the data. The model will be overly trained to the dataset, which may capture noise and create a non-generalized model. At the same time, conducting a train and test on the same dataset may result in a low out-of-sample accuracy because of over-fitting. A trained ANN needs to have a high out-of-sample accuracy since the goal is to make a correct prediction on unseen data.

In order to enhance the performance of the trained model on new data, this article employed the second approach (test the developed model on a set different from the training set). This study evaluates the model's performance using a test dataset of sample retrofit packages that the trained model has never used. The trained model accuracy is then evaluated by predicting those samples. The consistency between the predicted values and actual values on the test dataset serves as an indicator of the generalization efficiency of the ANN model.

2.1.4. Optimization of Hyperparameters

The hyperparameters of ANN can have an essential impact on the model's prediction performance. The hyperparameters mainly consist of the number of layers, number of neurons per layer, activation functions, dropout rate, optimizers, epochs, and batch size. However, most previous studies designed the structure and hyperparameters based on a trial-and-error approach and personal knowledge, where the developed ANN model may not achieve the best prediction performance. This research employed a genetic algorithm to optimize the model structure and thereby identify the best hyperparameters of the ANN model.

A genetic algorithm (GA) is an evolutionary algorithm utilized to address optimization problems in computational mathematics [18]. Evolutionary algorithms were originally developed from phenomena in evolutionary biology, including heredity, mutation, natural selection, and hybridization. For an optimization problem, a variety of candidate solutions can be abstractly represented as chromosomes so that the population evolves towards better solutions. This research adopted a well-designed GA tool to identify the best structure and hyperparameters of ANN [32]. The ANN-GA model implementation is described in Figure 3. The considered ANN hyperparameters and GA parameters are shown in Table 1.

2.2. Multi-Objective Optimization

Multi-objective optimization (MOO) was performed by coupling the well-trained ANN model with a Pareto optimization approach developed in the Python coding environment. The ANN model was employed to calculate the energy pre- and post-retrofit performance of the selected building. Predicted energy values were used to evaluate different retrofit strategies' emission and cost performances. Utilizing the Pareto optimization approach, the developed model can determine optimal solutions among various feasible retrofit strategies.

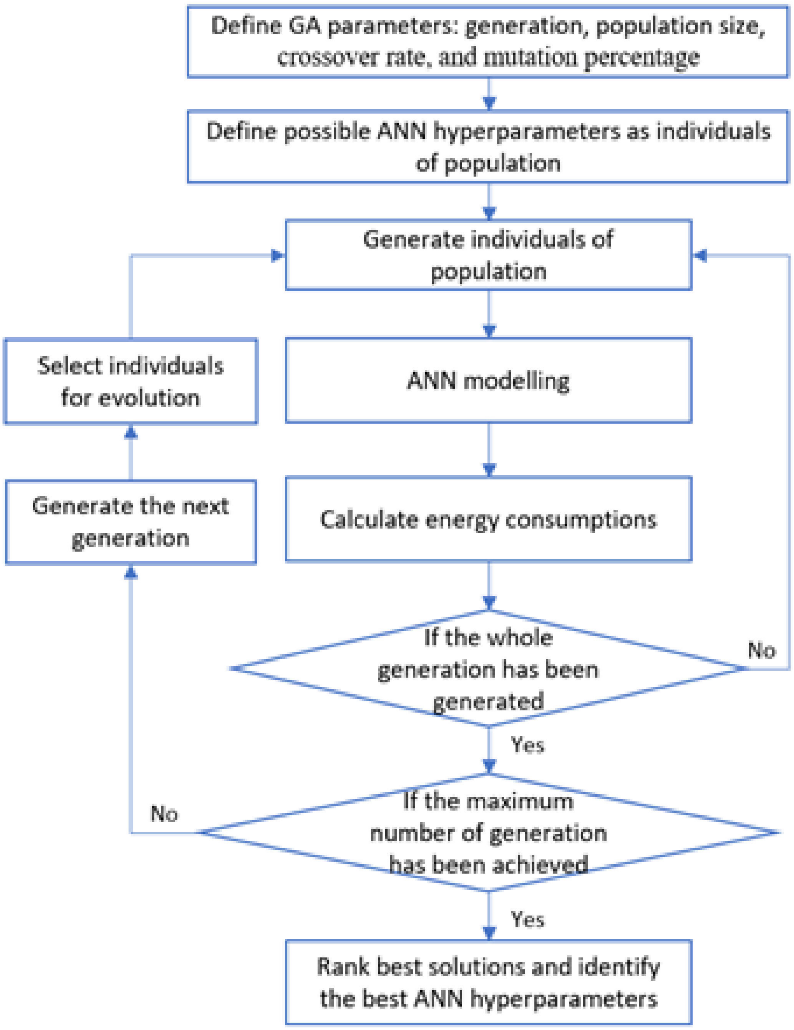


Figure 3. The implementation of ANN-GA model.

Table 1. ANN hyperparameters and GA parameters.

ANN Hyperparameters	GA Parameters
Number of layers: 2, 3, 4	Number of generations: 20
Number of neurons per layer: 1 to 20	Population size: 30
Activation functions: Relu, Sigmoid, Elu, and Tanh	Crossover rate: 20%
Optimizers: Adam, SGD, Adagrad, and RMSProp	Mutation percentage: 30%
Dropout rate: 0, 0.1, 0.2, 0.3, 0.4	
Epochs: 30, 50, 100, 200	
Batch size: 32, 64, 128, 256, 512	

2.2.1. Objective Functions

In Canada, conflicting stakeholder expectations toward environmental and economic performance enhancements are the main barrier to the implementation of retrofit programs [39]. For instance, public stakeholders (e.g., governments) pay more attention to reducing GHG emissions, while private stakeholders (e.g., homeowners) have more interest in saving utility bills through energy retrofits [5]. Merely accounting for emission reduction potential may place a heavy economic burden on private stakeholders. Thus, it is important to explore the trade-off between emission and cost savings and find an optimal balance between them. As such, the objective functions considered in this study are carbon emissions and retrofit costs, and the main objective is to minimize the two function values. The objective function values can be calculated by combining energy consumption results with emission and cost impact data.

Retrofit Emissions

The well-trained ANN model can calculate the annual natural gas and electricity consumption for a given building. The energy consumption data and emission factors were employed to determine the carbon emission for candidate retrofit solutions, as shown in Equation (5).

$$\Delta RE_{j,k} = \left(EE_k^{retrofitted} \times EF_e + NE_{j,k}^{retrofitted} \times EF_{NG}\right) \times T \tag{5}$$

where

- ΔRE_k = The carbon emission of the base building model by taking the k_{th} candidate solution (kg CO_2e).
- $EE_{j,k}^{retrofitted}$ = The electricity consumption of the building model by employing the k_{th} candidate solution (kWh).
- $NE_{j,k}^{retrofitted}$ = The natural gas energy consumption of the building model by taking the k_{th} candidate solution (GJ).
- EF_e = Local electricity grid emission factor (kg CO_2e/kWh).
- EF_{NG} = Natural gas emission factor (kg CO_2e/GJ).
- T = The remaining lifetime of the retrofitted building.

Retrofit Costs

The initial cost of a retrofit project is important for stakeholders. In Canada, householders are more interested in taking cheaper retrofit measures to reduce the initial cost [7]. However, taking a cheaper retrofit measure without accounting for energy cost savings may increase life cycle costs (LCC). The LCC accounts for all cost elements related to a retrofit project. A potential retrofit package with a higher initial cost might create a better LCC impact because of cost savings on utility bills. Life cycle cost assessment (LCCA), which can take into account immediate and long-term expenses, has been recognized as an essential economic evaluation method for building investments [40].

In terms of LCC calculations, the authors considered the initial retrofit cost and the operational energy cost. The end-of-life cost was ignored since reliable end-of-life cost data related to candidate retrofit solutions were not found in existing local databases. In addition, the end-of-life costs are usually dumped in the landfill or recycled by dealers without cost [7]. Therefore, end-of-life costs can be negligible compared to initial and operational costs.

The initial cost of a candidate retrofit solution refers to purchasing and installing building envelope components and heating and cooling energy equipment. The authors employed the RS Means 2019 database to calculate the initial costs of a variety of candidate retrofit solutions. The initial cost of a candidate retrofit solution can be determined by Equation (6).

$$IC_k = IC_{envelope} + IC_{energy\ system} = \sum_{i=1}^{a} A_i \times ic_i + \sum_{j=1}^{b} C_j \tag{6}$$

where

- IC_k = The initial retrofit cost of the k_{th} candidate retrofit solution.
- $IC_{envelope}$ = The initial costs of envelope component materials.
- $IC_{energy\ system}$ = The initial costs of energy equipment.
- cc_i = The unit initial cost of the i_{th} insulation material.

- C_j = The initial cost of the j_{th} energy equipment.

The operational cost consists of the energy cost savings and the annual maintenance cost of energy equipment. As shown in Equation (7), the annual operational cost savings can be determined by the energy consumption results produced by the well-trained ANN model and energy market prices.

$$\Delta AOC_k = \left(EE_{base} - EE_{retrofitted}\right) \times \mathrm{EP} + \left(NE_{base} - NE_{retrofitted}\right) \times \mathrm{NP} - \mathrm{MC} \quad (7)$$

where

- ΔAOC_k = The annual operational cost savings by taking the k_{th} candidate retrofit solution.
- EP = The local electricity price (CAD/kWh).
- NP = The local natural gas price (CAD/GJ).
- MC = The annual maintenance cost.

Considering the time value of money, the authors employed the net present value (NPV) of the operational cost savings in the calculation. The NPV of the operational energy cost savings can be identified by Equation (8).

$$\Delta OC_k = \sum_{t=0}^{t} \frac{\Delta AOC_k}{(1+r)^t} \quad (8)$$

where

- ΔOC_k = The net present value of operational energy cost savings.
- r = The discounted rate (%).
- t = The retrofit project lifetime.

The LCC of a candidate retrofit solution can be obtained by the following equation.

$$LCC_k = UC_k - \Delta OC_k \quad (9)$$

where

- LCC_k = The life cycle cost of the k_{th} candidate retrofit solution.

2.2.2. Pareto Optimization

The Pareto optimization approach is one of the most widely utilized MOO approaches for retrofit decision-making. Previous researchers have proved that this approach is robust when it is employed for a retrofit planning project [6,41,42]. MOO is effective when the issue is related to discontinuities that might exist in the output of machine-learning-based surrogate models because of discrete variables [21]. Candidate solutions in the "Pareto frontier" are regarded as optimal solutions for a given optimization problem. In this research, Pareto solutions suggest no other candidate retrofit packages that can improve one objective without reducing another objective. The development of the Pareto optimization approach is presented as follows:

1. Determine the inputs and the performance of the system's outputs (objective functions).
2. Produce all the candidate retrofit solutions and determine the objective function values for all the solutions, as presented in Figure 4a.
3. Determine the first front (F1) of non-dominated candidate retrofit solutions that other solutions do not dominate by using the approach, as shown in Figure 4b.
4. Exclude the candidate retrofit solutions of the first front and repeat Step (3) to obtain the second Pareto optimal solutions that are only dominated by the candidate retrofit solutions of the first front, as shown in Figure 4c.
5. Repeat Step (4)'s process until all candidate retrofit solutions belong to different fronts, as presented in Figure 4d.
6. Rank all the candidate retrofit solutions based on the order of fronts.

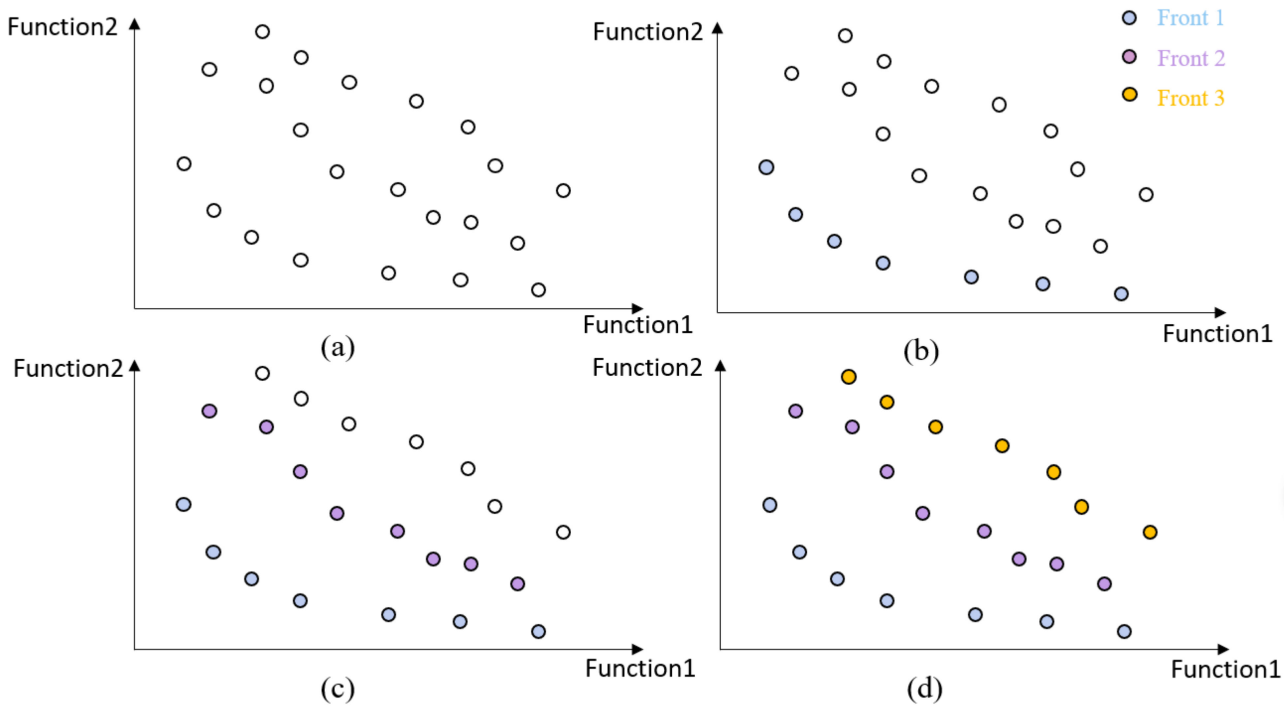


Figure 4. Pareto optimization approach.

2.3. Technique for Order of Preference by Similarity to Ideal Solution

After identifying Pareto optimal retrofit solutions, this article employed the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to conduct the relative performance evaluation of these optimal solutions and select the best retrofit package. In general, TOPSIS is a multi-criteria decision-making (MCDM) method of compensatory aggregation that compares a limited number of evaluation objectives by determining weights for each criterion, normalizing scores for each criterion, and calculating the geometric distance between each evaluation objective and the ideal target [43]. The method is employed to evaluate, rank, and compare retrofit packages with the selected criteria and indicators. The procedures for applying the TOPSIS method are described as follows:

(1) Weighted values of indicators

The weighted value of the selected indicator is determined by the following equation.

$$v_{ij} = w_j * r_{ij} \tag{10}$$

where v_{ij} is the weighted value of the j_{th} indicator for the i_{th} retrofit solution, w_j is the weight of the j_{th} indicator, and r_{ij} is the original value of the j_{th} indicator for the i_{th} retrofit solution.

(2) The positive and negative ideal solutions

The positive ideal solution (A^+) is a composite of the best performance values of a sample solution across all criteria, while the negative ideal solution (A^-) is a composite of the worst performance values. The two solutions can be determined by the following equations:

$$A^+ = \{v_1^+, v_2^+ \ldots v_n^+\} \tag{11}$$

$$A^- = \{v_1^-, v_2^- \ldots v_n^-\} \tag{12}$$

where

$$v_j^+ = \begin{cases} \max_i\{v_{ij}\}, \text{ if } j \text{ is a positive criteria} \\ \min_i\{v_{ij}\}, \text{ if } j \text{ is a negative criteria} \end{cases} \tag{13}$$

$$v_j^- = \begin{cases} \max_i\{v_{ij}\}, & \text{if } j \text{ is a negative criteria} \\ \min_i\{v_{ij}\}, & \text{if } j \text{ is a positive criteria} \end{cases} \quad (14)$$

(3) Separation measures

The Euclidean distances from a possible retrofit solution *i* to the positive ideal solution (S^+) and negative ideal solution (S^-) are calculated by the following equations:

$$S_i^+ = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^+\right)^2}, \ i = 1, \ldots m \quad (15)$$

$$S_i^- = \sqrt{\sum_{j=1}^{m} \left(v_{ij} - v_j^-\right)^2}, \ i = 1, \ldots m \quad (16)$$

(4) Similarities to the positive ideal solution

The closeness coefficient (C_i^*) is utilized to represent the similarities between a candidate retrofit solution to the negative ideal solution. A larger value of closeness indicates better performance of a sample renewal retrofit project because a longer distance to the negative ideal solution presents a better evaluation result.

$$C_i^* = \frac{S_i^-}{S_i^+ + S_i^-} \quad (17)$$

The final score of retrofit solutions and their rankings is based on their performance level. Following the TOPSIS procedures, the retrofit solutions can be ranked in descending order according to their closeness coefficients.

3. Case Study

In order to demonstrate the proposed framework, a case study was conducted for a dataset that records the energy consumption of retrofit scenarios for a reference building located in British Columbia, Canada. The reference building is a medium single-family detached house with a shallow basement. The building characteristic and investigated energy retrofit options are shown in Table 2. The representative dataset of a wide range of retrofit scenarios was produced using building energy modelling tools (HOT2000 and HTAP). The details of the energy modelling process were described in the research [6]. This dataset consists of 10,368 retrofit scenarios and the corresponding natural gas and electricity consumption values [44].

Table 2. Investigated retrofit options for inputs.

Components	Building Characteristics	Considered Retrofit Options
Space heating equipment	Natural gas forced air furnace with 65% annual fuel utilization efficiency (AFUE)	Natural gas furnace (92%AFUE), electric baseboard and heat pump
Water heating equipment	Natural gas conventional tank	Natural gas instantaneous, conventional electric tank, electric heat pump
Airtightness	12 Air Change per Hour (ACH) 50 Pa	5 ACH 50 Pa, 3.5 ACH 50 Pa
Above-grade wall insulation	Thermal resistance value R8.49	Thermal resistance value R17.5, R22, R28, R30, R38 blown-in fibreglass insulation
Below-grade wall insulation	Thermal resistance value R3.24	Thermal resistance value R17.5 blown-in fibreglass insulation
Ceiling insulation	Thermal resistance value R8.47	Thermal resistance value R40, R50, R60 batt insulation
Windows	Single-glazed wood frame	Double-pane frame, triple-pane frame
Exterior door	Solid wood	Fibreglass polystyrene core

The dataset was utilized for training and validating ANN models and thereby determining the optimal model structure and hyperparameters using the aforementioned genetic algorithm. ANN models were trained using a personal computer with an i7-8750H processor, 2.2 GHz CPU, and 16 GB RAM. A total of 70% of the data were used to train the models, while 20% of the data were used for validation to tune the hyperparameters of the models. Finally, the remaining 10% of the data were considered as the test data to evaluate the generalization efficiency of the trained models. The optimization and training process was implemented in the Python coding environment by employing TensorFlow, Scikit-learn, Numpy, and Pandas libraries.

After identifying the optimal ANN model structure and hyperparameters, the study developed a well-trained ANN model to calculate natural gas and electricity consumption. By combining emission and cost impact data, the two indicators were used to determine carbon emissions and energy costs, considering objective functions for the following multi-objective optimization. Pareto optimal retrofit solutions can be identified by using the approach mentioned above. Finally, the TOPSIS method was employed to rank the identified Pareto optimal solutions. In this study, the weight of each objective was assumed to be equal (0.5). Decision-makers can adjust the weights according to their needs and budgets and thus select the best retrofit scenario.

4. Results and Discussion

The developed ANN model was applied to the data of the case study to evaluate the energy performance of different retrofit packages for a given building. In this section, the training results, as well as the prediction performance of the ANN model for the test dataset, are shown and discussed. In addition, Pareto optimal retrofit solutions are identified while considering the emission and cost performance using a multi-objective optimization method. Finally, prioritized retrofit packages are determined based on the aforementioned MCDM approach.

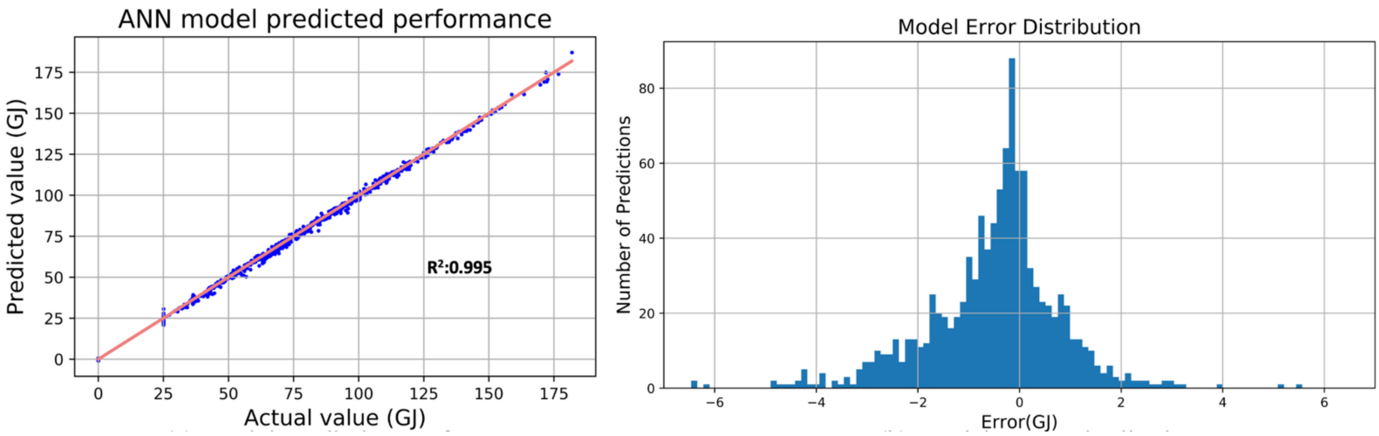
4.1. ANN Model Prediction Performance

Among the different hyperparameters tested, the best performance was achieved with the following configuration:

- ANN with eight input variables and two output variables.
- MLP with four hidden layers of neurons: [15–18].
- Activation function: Relu function.
- Optimizer: Adam.
- Dropout rate: 0.1.
- Batch size: 64.
- Epochs: 100.

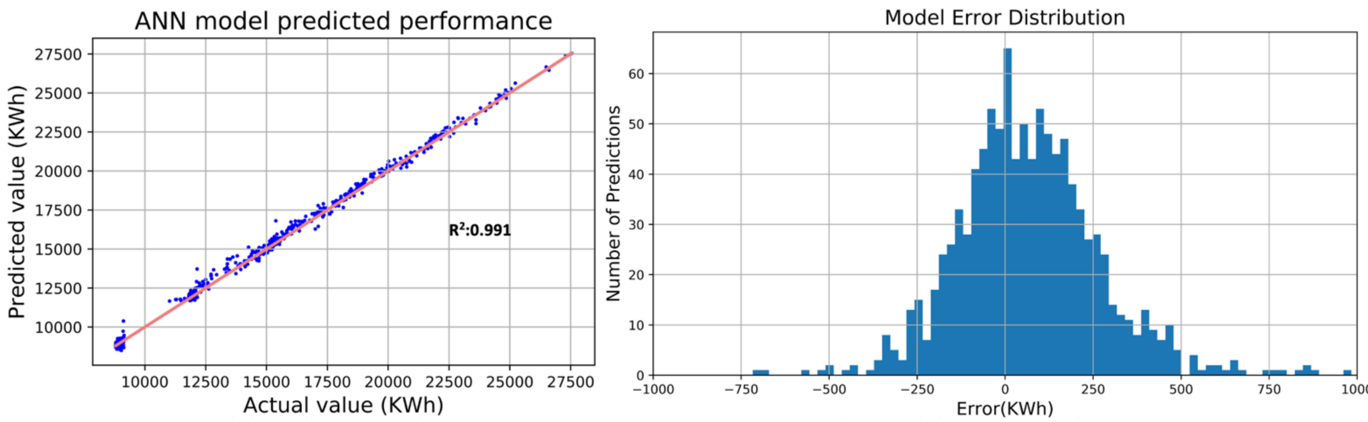
The sample dataset is categorized into three groups: the training, the derivative, and the test datasets. The training and derivate dataset were utilized to ensure that the trained model is not overfitting or underfitting. The datasets were used to confirm the model's generalization and tune the model's hyperparameters. The test dataset was used to present the generalization ability and explain the prediction performance in an unseen dataset.

The prediction performance of energy consumption on the test set of retrofit packages is presented in Figures 5 and 6. As shown in Table 3, the performance of the condition indicators is measured with the use of R^2, RMSE, and MAE. The test results present a good prediction performance for the annual NG consumption and electricity consumption, with an average coefficient of determination score of R^2 = 0.995 and 0.991, respectively. A root mean squared error of 1.384 GJ is achieved for the ANN prediction for the NG consumption, while an absolute average error of 0.981 GJ is observed. For the annual electricity consumption, the overall average prediction performance of the objective values are RMSE = 226.023 KWh and MAE = 163.552 KWh. After calculating the annual NG and electricity consumption, the carbon emissions and life cycle costs can be identified. The two indicators are considered objective functions for the following optimization.



(**a**) (**b**)

Figure 5. The prediction performance of annual NG consumption: (**a**) Model Prediction Performance; (**b**) Model Error Distribution.



(**a**) (**b**)

Figure 6. The prediction performance of annual electricity consumption: (**a**) Model Prediction Performance; (**b**) Model Error Performance.

Table 3. ANN model performance.

Energy Performance	RMSE	MAE	R^2
NG consumptions (GJ)	1.384	0.981	0.995
Electricity consumptions (KWh)	226.023	163.552	0.991

4.2. Pareto Optimal Retrofit Solutions

According to the above-mentioned analysis, Pareto optimal retrofit solutions can be identified, as shown in Figure 7. The grey points represent possible retrofit packages, while the red points represent Pareto optimal solutions. Decision-makers can choose different retrofit solutions according to their budgets and emission reduction target. As shown in the figure, the solutions located in the blue oval are regarded as primary solutions, which can reduce carbon emissions and achieve a negative life cycle cost. This means that the energy cost savings are larger than initial retrofit costs. In practice, governments can provide awareness information programs to help occupants learn about the cost-saving benefits of retrofits. However, other Pareto solutions are recognized as secondary solutions. These solutions can reduce more carbon emissions but need more investments. Governments need to provide financial support for occupants, such as grants and rebates, to reduce their economic burdens and encourage them to implement energy retrofits.

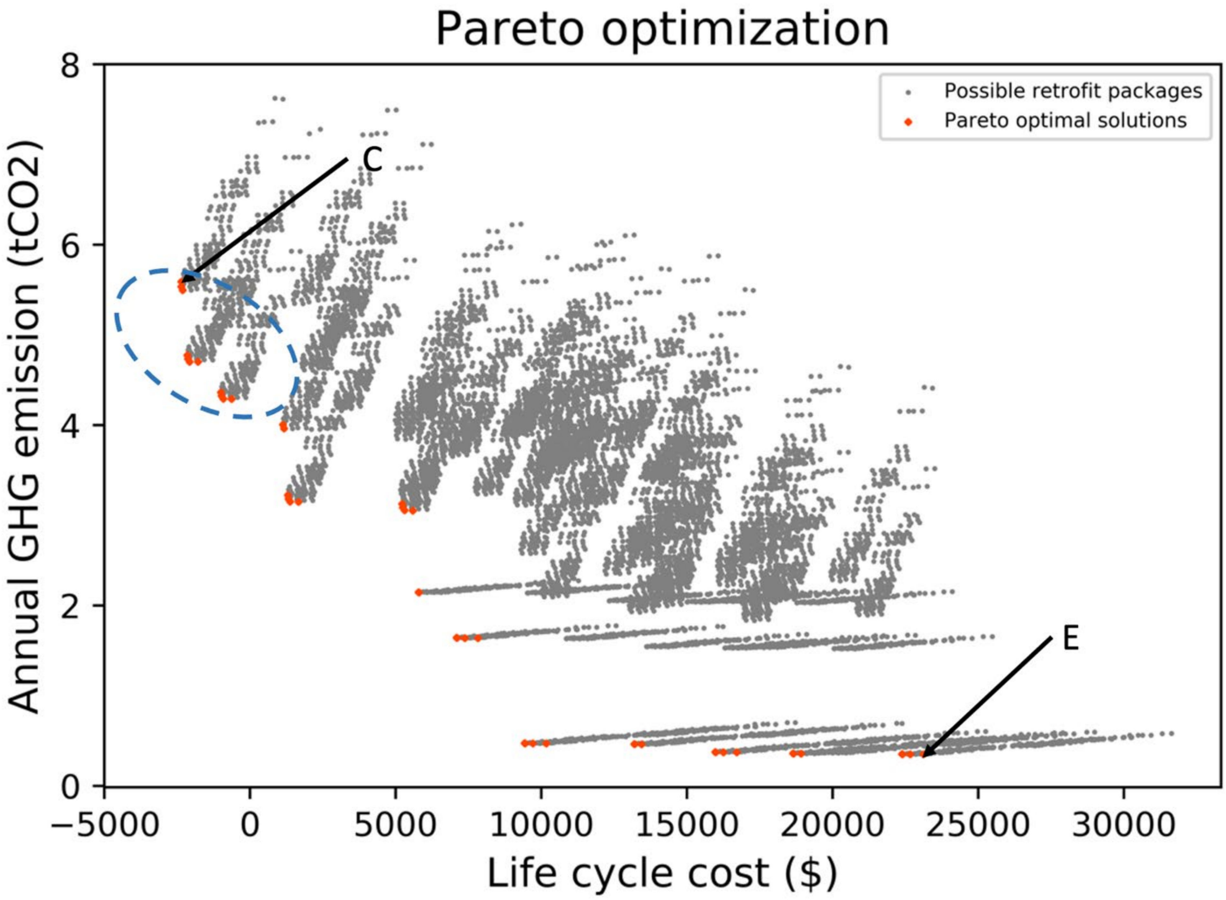


Figure 7. Pareto optimal retrofit solutions.

There are two special Pareto solutions highlighted in the figure among these solutions. The first one (point C) is named the "cost-optimal solution (COS)", which can activate the most cost savings (the lowest LCC). This solution can assist occupants in saving utility bills most effectively but can produce a relatively high carbon emission. The second one (point E) is termed the "emission-optimal solution (EOS)", which can help governments achieve the most environmental benefits with a high investment. While the two unique solutions are meaningful for private and public stakeholders, it is essential to explore the trade-off between emission reductions and cost savings to find an optimal balance. As such, this study employed multi-criteria decision-making analysis to assist decision-makers in choosing the best retrofit solution according to their needs.

4.3. Prioritized Retrofit Packages

The two-dimensional Euclidean distance of each optimal retrofit solution was calculated using the multi-criteria decision-making method. The first step was to calculate the weighted values v_{ij}, determined by normalizing and weighting the Pareto solutions. The positive and negative ideal solutions on the Pareto-front set can be identified, as shown in Figure 8. Then, the Euclidean distance from each solution to the positive ideal solution (S_i^+) and negative ideal solution (S_i^-) and their relative closeness to the ideal solution C_i^*, termed TOPSIS scores, can be calculated. All solutions in the Pareto-front set are ranked according to TOPSIS scores. The point with the maximum TOPSIS score was identified to be the best retrofit package.

With the multi-objective optimization and multi-criteria decision-making process, the top five retrofit packages and the corresponding cost savings (percentages of savings in energy consumption) and emission reductions can be identified, as shown in Table 4. Envelope upgrades, including below-grade wall insulation (R17.5), above-grade wall insulation, and ceiling insulation, are the most common retrofit options derived from the TOPSIS analysis. In addition, airtightness enhancement (3.5 ACH) and an electric heat pump for water heating are also recommended. Implementing the retrofit packages can produce cost savings of over CAD 2000 and emission reductions of over 1.7 tCO_2, indicating that these packages are primary retrofit solutions, which can produce both environmental and economic benefits.

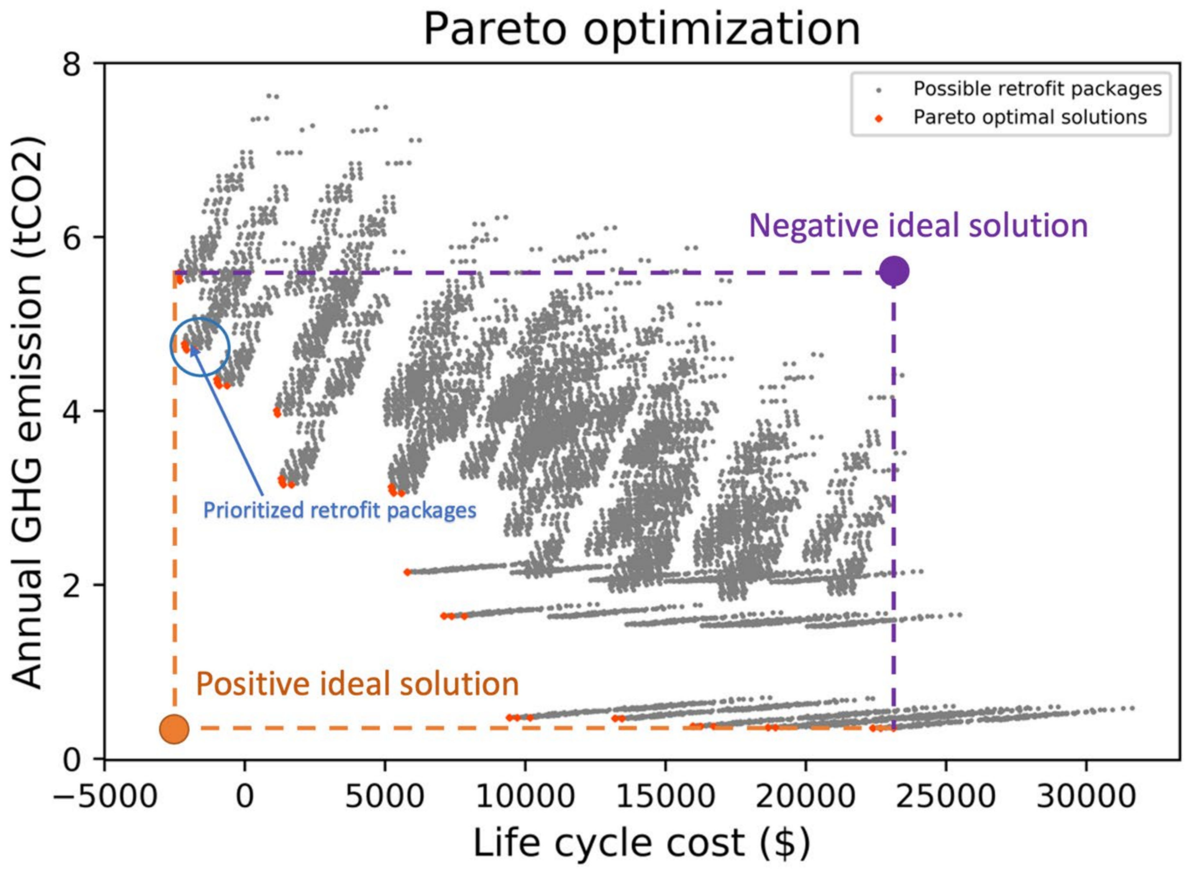


Figure 8. Prioritized retrofit packages on the Pareto-front set.

Table 4. Prioritized retrofit packages.

Rank	Prioritized Retrofit Packages	Cost Savings (Percentages)	Emission Reductions (tCO_2)
1	Airtightness: 3.5 ACH, below-grade wall insulation: R17.5, water heating equipment: electric heat pump	2.3 K (75%)	1.76
2	Airtightness: 3.5 ACH, above-grade wall insulation: R28, below-grade wall insulation: R17.5, ceiling insulation: R60, space heating equipment: natural gas furnace	2.3 K (73%)	1.80
3	Above-grade wall insulation: R22, ceiling insulation: R60, water heating equipment: electric heat pump	2.3 K (70%)	1.84
4	Above-grade wall insulation: R38, ceiling insulation: R40, water heating equipment: electric heat pump	2.1 K (48%)	2.58
5	Airtightness: 3.5 ACH, above-grade wall insulation: R22, below-grade wall insulation: R17.5, ceiling insulation: R40, water heating equipment: natural gas, instantaneous	2.1 K (47%)	2.62

As retrofit activities require economic support, financial incentives have garnered greater attention when implementing energy retrofit projects. Financial incentives can address stakeholders' main concerns on economic issues related to retrofit projects, such as high initial cost and long payback period, and thus improve their willingness to renovate their buildings [45]. However, financial incentives might impose a heavy financial burden on governments in the long run [46]. According to the analysis mentioned above, the identified optimal retrofit packages can produce both environmental and economic benefits. This means that governments can provide more awareness information programs for occupants to help them understand the economic benefits of retrofits and thereby encourage them to uptake retrofits. The awareness information programs can help in reducing the financial burden on governments when implementing retrofit schemes.

5. Conclusions

Building energy retrofits have been recognized as key to reducing building energy consumption and associated GHG emissions in Canada. This paper proposed a data-driven framework that combines machine learning, MOO, and MCDM techniques to evaluate energy consumption, GHG emissions, and energy costs of a wide range of retrofit scenarios and formulate optimal retrofit plans. First, the authors developed a well-trained ANN model as a surrogate tool to imitate the building energy modelling process, where the optimal model hyperparameters were identified using a genetic algorithm. Natural gas and electricity consumption can be calculated using the ANN model. Then, the authors calculated carbon emissions and energy costs of retrofit scenarios by integrating energy consumption with emission and cost impact data, respectively. The calculated emissions and costs were considered objective functions for the following multi-objective optimization to determine Pareto optimal retrofit solutions. Finally, the authors identified the best retrofit scenario using an MCDM approach.

The proposed framework was validated using an existing dataset for a reference building located in British Columbia, Canada. For the trained ANN model, the overall average prediction performance is observed for the natural gas consumption and electricity consumption with average values of R^2 = 0.995 and R^2 = 0.991, respectively. According to the optimization results, the most recommended retrofit package includes the upgrades in airtightness (3.5 ACH), below-grade wall insulation (R17.5), and water heating equipment (electric heat pump), which can produce both environmental (over 1.7 tCO_2 reductions) and economic benefits (over CAD 2000 cost savings) for retrofit stakeholders. The case study proved that the proposed framework could effectively predict the energy consumption of a wide range of retrofit packages and help decision-makers make an optimal decision on developing retrofit plans.

The proposed framework can be modified to accustom different building types in other provinces across Canada. However, some limitations exist in this study that the authors will address in future work. This study employed an artificial neural network to predict building energy consumption. Further studies need to be conducted to compare the prediction accuracy of other machine learning models, such as support vector machines and gradient boosting. In addition, the input features can be improved by adding more variables, such as weather data and climate zones.

Author Contributions: Conceptualization, H.Z.; methodology, H.Z.; software, H.Z.; validation, H.Z.; formal analysis, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, H.F., K.H. and M.A.; supervision, K.H. and M.A.; funding acquisition, H.F., K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MITACS and FortisBC.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to gratefully acknowledge MITACS and FortisBC for funding and supporting this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Karunathilake, H.; Hewage, K.; Sadiq, R. Opportunities and challenges in energy demand reduction for Canadian residential sector: A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2005–2016. [CrossRef]
2. Feng, H.; Zhao, J.; Zhang, H.; Zhu, S.; Li, D. Uncertainties in whole-building life cycle assessment: A systematic review. *J. Build. Eng.* **2022**, *50*, 104191. [CrossRef]
3. Natural Resources Canada. Energy and Greenhouse Gas Emissions. 2018. Available online: https://www.nrcan.gc.ca/science-data/data-analysis/energy-data-analysis/energy-facts/energy-and-greenhouse-gas-emissions-ghgs/20063 (accessed on 24 March 2022).
4. Government of British Columbia. How It Works: Energy Step Code. 2018. Available online: https://energystepcode.ca/how-it-works/ (accessed on 24 March 2022).
5. Prabatha, T.; Hewage, K.; Karunathilake, H.; Sadiq, R. To retrofit or not? Making energy retrofit decisions through life cycle thinking for Canadian residences. *Energy Build.* **2020**, *226*, 110393. [CrossRef]
6. Zhang, H.; Hewage, K.; Prabatha, T.; Sadiq, R. Life cycle thinking-based energy retrofits evaluation framework for Canadian residences: A Pareto optimization approach. *Build. Environ.* **2021**, *204*, 108115. [CrossRef]
7. Feng, H.; Liyanage, D.R.; Karunathilake, H.; Sadiq, R.; Hewage, K. BIM-based life cycle environmental performance assessment of single-family houses: Renovation and reconstruction strategies for aging building stock in British Columbia. *J. Clean. Prod.* **2020**, *250*, 119543. [CrossRef]
8. Bakhtavar, E.; Prabatha, T.; Karunathilake, H.; Sadiq, R.; Hewage, K. Assessment of renewable energy-based strategies for net-zero energy communities: A planning model using multi-objective goal programming. *J. Clean. Prod.* **2020**, *272*, 122886. [CrossRef]
9. Karunathilake, H.; Hewage, K.; Brinkerhoff, J.; Sadiq, R. Optimal renewable energy supply choices for net-zero ready buildings: A life cycle thinking approach under uncertainty. *Energy Build.* **2019**, *201*, 70–89. [CrossRef]
10. Ma, Z.J.; Cooper, P.; Daly, D.; Ledo, L. Existing building retrofits: Methodology and state-of-the-art. *Energy Build.* **2012**, *55*, 889–902. [CrossRef]
11. Ascione, F.; Bianco, N.; de Stasio, C.; Mauro, G.M.; Vanoli, G.P. Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy* **2017**, *118*, 999–1017. [CrossRef]
12. Pilechiha, P.; Mahdavinejad, M.; Rahimian, F.P.; Carnemolla, P.; Seyedzadeh, S. Multi-objective optimisation framework for designing office windows: Quality of view, daylight and energy efficiency. *Appl. Energy* **2020**, *261*, 114356. [CrossRef]
13. Sharif, S.A.; Hammad, A. Simulation-Based Multi-Objective Optimization of institutional building renovation considering energy consumption, Life-Cycle Cost and Life-Cycle Assessment. *J. Build. Eng.* **2019**, *21*, 429–445. [CrossRef]
14. Dell'Anna, F.; Bottero, M.; Becchio, C.; Corgnati, S.P.; Mondini, G. Designing a decision support system to evaluate the environmental and extra-economic performances of a nearly zero-energy building, Smart Sustain. *Built Environ.* **2020**, *9*, 413–442. [CrossRef]
15. Broday, E.E.; da Silva, M.C.G. The role of internet of things (IoT) in the assessment and communication of indoor environmental quality (IEQ) in buildings: A review, Smart Sustain. *Built Environ.* **2022**. [CrossRef]
16. Birgonul, Z. A receptive-responsive tool for customizing occupant's thermal comfort and maximizing energy efficiency by blending BIM data with real-time information, Smart Sustain. *Built Environ.* **2021**, *10*, 504–535. [CrossRef]
17. Tronchin, L.; Manfren, M.; Tagliabue, L.C. Optimization of building energy performance by means of multi-scale analysis—Lessons learned from case studies. *Sustain. Cities Soc.* **2016**, *27*, 296–306. [CrossRef]
18. Ascione, F.; Bianco, N.; de Stasio, C.; Mauro, G.M.; Vanoli, G.P. A new methodology for cost-optimal analysis by means of the multi-objective optimization of building energy performance. *Energy Build.* **2015**, *88*, 78–90. [CrossRef]
19. Thrampoulidis, E.; Mavromatidis, G.; Lucchi, A.; Orehounig, K. A machine learning-based surrogate model to approximate optimal building retrofit solutions. *Appl. Energy* **2021**, *281*, 116024. [CrossRef]
20. Hong, T.; Wang, Z.; Luo, X.; Zhang, W. State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build.* **2020**, *212*, 109831. [CrossRef]
21. Cecconi, F.R.; Moretti, N.; Tagliabue, L.C. Application of artificial neutral network and geographic information system to evaluate retrofit potential in public school buildings, Renew. Sustain. *Energy Rev.* **2019**, *110*, 266–277. [CrossRef]
22. Grillone, B.; Danov, S.; Sumper, A.; Cipriano, J.; Mor, G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110027. [CrossRef]
23. Seyedzadeh, S.; Rahimian, F.P.; Oliver, S.; Rodriguez, S.; Glesk, I. Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making. *Appl. Energy* **2020**, *279*, 115908. [CrossRef]
24. Miller, C.; Nagy, Z.; Schlueter, A. A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1365–1377. [CrossRef]
25. Maddalena, E.T.; Lian, Y.; Jones, C.N. Data-driven methods for building control—A review and promising future directions. *Control Eng. Pract.* **2020**, *95*, 104211. [CrossRef]
26. D'Amico, A.; Ciulla, G.; Traverso, M.; Brano, V.L.; Palumbo, E. Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study. *J. Clean. Prod.* **2019**, *239*, 117993. [CrossRef]

27. Beccali, M.; Ciulla, G.; Brano, V.L.; Galatioto, A.; Bonomolo, M. Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in Southern Italy. *Energy* **2017**, *137*, 1201–1218. [CrossRef]
28. Seyedzadeh, S.; Rahimian, F.P.; Oliver, S.; Glesk, I.; Kumar, B. Data driven model improved by multi-objective optimisation for prediction of building energy loads. *Autom. Constr.* **2020**, *116*, 103188. [CrossRef]
29. Seyedzadeh, S.; Rahimian, F.P.; Rastogi, P.; Glesk, I. Tuning machine learning models for prediction of building energy loads. *Sustain. Cities Soc.* **2019**, *47*, 101484. [CrossRef]
30. Martins, L.A.; Soebarto, V.; Williamson, T.; Pisaniello, D. Personal thermal comfort models: A deep learning approach for predicting older people's thermal preference. *Smart Sustain. Built Environ.* **2022**. [CrossRef]
31. Francis, A.; Thomas, A. System dynamics modelling coupled with multi-criteria decision-making (MCDM) for sustainability-related policy analysis and decision-making in the built environment. *Smart Sustain. Built Environ.* **2022**. [CrossRef]
32. Kerdan, I.G.; Gálvez, D.M. Artificial neural network structure optimisation for accurately prediction of exergy, comfort and life cycle cost performance of a low energy building. *Appl. Energy* **2020**, *280*, 115862. [CrossRef]
33. Hüthwohl, P.; Lu, R.; Brilakis, I. Multi-classifier for reinforced concrete bridge defects. *Autom. Constr.* **2019**, *105*, 102824. [CrossRef]
34. Kaab, A.; Sharifi, M.; Mobli, H.; Nabavi-Pelesaraei, A.; Chau, K.W. Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production. *Sci. Total Environ.* **2019**, *664*, 1005–1019. [CrossRef] [PubMed]
35. Ouache, R.; Nahiduzzaman, K.M.; Hewage, K.; Sadiq, R. Performance Investigation of Fire Protection and Intervention Strategies: Artificial Neural Network-based Assessment Framework. *J. Build. Eng.* **2021**, *42*, 102439. [CrossRef]
36. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
37. Ciresan, D.; Meier, U.; Schmidhuber, J. Multi-column Deep Neural Networks for Image Classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3642–3649. [CrossRef]
38. Le, Q.V. Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8595–8598. [CrossRef]
39. Zhang, H.; Hewage, K.; Karunathilake, H.; Feng, H.; Sadiq, R. Research on policy strategies for implementing energy retrofits in the residential buildings. *J. Build. Eng.* **2021**, *43*, 103161. [CrossRef]
40. *ISO 15686-5:2017(en)*; Buildings and Constructed Assets—Service Life Planning—Part 5: Life-Cycle Costing. International Organization for Standardization: Geneva, Switzerland, 2017.
41. Delmastro, C.; Mutani, G.; Corgnati, S.P. A supporting method for selecting cost-optimal energy retrofit policies for residential buildings at the urban scale. *Energy Policy* **2016**, *99*, 42–56. [CrossRef]
42. Shaikh, P.H.; Shaikh, F.; Sahito, A.A.; Uqaili, M.A.; Umrani, Z. An Overview of the Challenges for Cost-Effective and Energy-Efficient Retrofits of the Existing Building Stock. In *Cost-Effective Energy Efficient Building Retrofitting:Materials, Technologies, Optimization and Case Studies*; Elsevier: Amsterdam, The Netherlands, 2017. [CrossRef]
43. Zavadskas, E.K.; Zakarevicius, A.; Antucheviciene, J. Evaluation of Ranking Accuracy in Multi-Criteria Decisions. *Informatica* **2006**, *17*, 601–618. [CrossRef]
44. MyronZhang13/MachineLearning_Python: This Is a Repository Used for the Machine Learning with Python Course. Available online: https://github.com/MyronZhang13/MachineLearning_Python (accessed on 19 April 2022).
45. Galvin, R.; Sunikka-Blank, M. Economic viability in thermal retrofit policies: Learning from ten years of experience in Germany. *Energy Policy* **2013**, *54*, 343–351. [CrossRef]
46. Rana, A.; Sadiq, R.; Alam, M.S.; Karunathilake, H.; Hewage, K. Evaluation of financial incentives for green buildings in Canadian landscape. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110199. [CrossRef]