Journal of Building Engineering 111 (2025) 113341

Contents lists available at ScienceDirect

Journal of Building Engineering

journal homepage: www.elsevier.com/locate/jobe

Transfer learning for cross-building forecasting of building energy and indoor air temperature in model predictive control applications

Hongwen Dou *, Kun Zhang

Department of Mechanical Engineering, École de technologie supérieure, Montreal, Canada

ARTICLE INFO

Keywords:
Transfer learning
Neural network model
Model predictive control
Building energy
HVAC

ABSTRACT

When applying Model Predictive Control (MPC) for Heating, Ventilation and Air Conditioning (HVAC) systems in buildings, accurate forecasting of short-term energy demand and indoor air condition profiles is essential. However, new or retrofitted buildings lack sufficient operation data to develop precise data-driven models. This study investigates transfer learning techniques to enhance the forecasting performance of black-box models under limited data conditions. Specifically, we leverage synthetic data from an open-source EnergyPlus building model to pre-train three neural network models, which are then transferred to a real building and fine-tuned with limited measurements. The results indicate that incorporating synthetic data into the pre-training phase significantly enhances the forecasting accuracy for building and HVAC energy, as well as indoor air temperature profiles, over a 12-h horizon with 15-min intervals. The study underscores the potential of combining transfer learning with synthetic data to address data limitations, extending the applicability of learning-based MPC in real-world buildings.

1. Introduction

Buildings account for a significant portion of energy consumption worldwide, while in Canada, the building sector represents about 30 % of national energy use [1]. In the context of grid-interactive efficient buildings, building energy demand predictions are an essential step in bringing that vision into fruition, for example, when integrating smart technologies into buildings, such as Model Predictive Control (MPC), data analytics, and automatic fault detection and diagnostics for energy efficiency and demand flexibility [2, 3]. In the case of applying MPC on HVAC systems, accurate forecasting of indoor air conditions is also critical as it ensures that the MPC controller can achieve optimal energy performance without compromising thermal comfort for the occupants [4–8]. However, accurately forecasting short-term profiles of building and HVAC system energy demand or indoor air conditions remains challenging due to continuous external and internal disturbances in buildings, such as changing weather conditions, occupancy, plug loads, etc. [9, 10].

Recent advancements in the domain of machine learning and artificial intelligence have facilitated the application of black-box models in building-related forecasting or predictions. Despite their potential, these types of models generally require a substantial number of datasets for training and calibration to achieve acceptable precision. In a real-world setting, the Building Automation System (BAS) in existing buildings is typically a valuable data resource for model development. However, security or privacy concerns

* Corresponding author.
E-mail address: hongwen.dou@etsmtl.ca (H. Dou).

https://doi.org/10.1016/j.jobe.2025.113341
Received 26 March 2025; Received in revised form 19 June 2025; Accepted 26 June 2025
Available online 27 June 2025

could limit access to or usage of the data. Even if all BAS data were accessible, it was likely that the available data could not be directly used for the modelling purpose, as it is quite common that the BAS data contains noise, errors, or gaps. More commonly, necessary variables required for the specific project objective may not be measured. When the project is a new or retrofitted building, or newly equipped with a BAS system, the operation data is lacking.

To tackle the problem of little, poor or no historical data as discussed above, various methods have been utilized to augment datasets. One such method involves integrating sophisticated mathematical models, such as convolutional neural network models [11, 12]. However, the data generated by these methods might be less representative of real-world situations and difficult to extrapolate, compared with synthetic data generated by physics-based models.

Another promising approach is Transfer Learning (TL), which involves transferring knowledge from a source domain with abundant data to a target domain with limited data through machine learning models. In other words, TL allows models to leverage knowledge gained from one domain (a building in our case) to improve performance in another, making it particularly useful in cases where data from the target environment is scarce. Many studies have focused on transfer learning to transfer knowledge from a source building to improve forecasting performance in a target building, typically assuming that both buildings share the same type of data. However, the question remains: how does transfer learning perform when the source building uses synthetic data while the target building relies on measured data? In this research, we utilize synthetic data generated from an open-source EnergyPlus [13] (a physics-based building energy simulation program) model from the U.S. Department of Energy (DOE) prototype buildings [14] to pre-train neural network models. These pre-trained models are then fine-tuned using a limited set of real measurements from an actual building, i.e., the target building. The real and virtual buildings, though sharing the same architectural type (i.e., the large commercial building), are located in different cities, each subject to unique weather and environmental conditions.

We then explore two distinct TL strategies and evaluate their effectiveness using three neural network models. By comparing the results, we aim to determine how well TL techniques can improve the accuracy of the forecasting models. Specifically, we investigate how well the models forecast building and HVAC energy demand and indoor air temperature profiles over a 12-h horizon at 15-min intervals, as these three variables are crucial when implementing MPC on HVAC systems. The performance of these models is then evaluated in two dimensions: the absolute performance according to the Coefficient of Variation of the Root Mean Squared Errors (CVRMSE), and the relative performance against a baseline, in which the TL techniques are not used, and neural network models are generated using the limited data from the target building. With the verified forecasting performance of the models, we envision their future integration into a supervisory MPC controller. Such a controller could send optimal indoor air temperature setpoints to thermostats to achieve defined project goals, such as activating building energy flexibility.

The paper is structured as follows: Section 2 presents the literature review, research gaps and contributions of this paper. Section 3 details the methodology of the TL strategies, the three neural network models and the performance indicators used to evaluate the models. Section 4 presents details of the case study buildings: the source and target buildings. Section 5 discusses and analyzes the simulation results, followed by the conclusion and recommendations in Section 6.

2. Background and contribution

2.1. Related works

TL enables the transfer of knowledge from one domain to another, allowing models to extrapolate beyond their original scope. It involves two primary components: feature and marginal space [15]. The feature space refers to the input variables used for learning (e.g., indoor air temperature, outdoor air temperature, or HVAC loads), while the marginal space refers to the distribution of these features (i.e., how frequently or in what patterns different values appear). These two spaces can differ between the source and target domains, either in terms of their feature spaces, marginal spaces, or both. Another key component of TL is the task, which consists of a label space and a function used to estimate unseen instances. Typically, the source task represents the learning task within the source domain, while the target task represents the learning task within the target domain. Yang et al. [15] formally defined TL as:

> "Given a source domain and a learning task, and a target domain and a learning task, the transfer learning aims to help improve the learning of the target predictive function $f_t(\cdot)$ for the target domain using the knowledge in the source domain and the source task, where the source domain does not equal to the target domain or the source task does not equal to the target task."

TL can be categorized as either homogeneous or heterogeneous, depending on the similarity of the feature spaces and target spaces between the source and the target domain [16,17]. Click or tap here to enter text.When the features or labels differ between a source domain and a target domain, it constitutes heterogeneous TL; otherwise, it is considered homogeneous TL.

TL has been successfully applied in various fields, such as improving the deep learning model accuracy and stability of the photovoltaic power system [18], the fault diagnosis of the oil-gas treatment station [19], and imagine classification [20] etc. In the field of building energy forecasting or prediction, TL was usually used to transfer Building Automation System (BAS) data from source buildings to target buildings that also have BAS data. When both source buildings and target buildings have measurement data for TL tasks, they benefit from more relevant and similar data distributions. Thus, this application scenario leverages the real-world data for both training and fine-tuning, improving overall model performance. A few examples are listed herein for such scenarios.

In a study focusing on 1-h ahead forecasting of building energy consumption, both the source and target buildings utilized real measurement data [21]. Among the 40 buildings with diverse functions, ranging from office and residential to commercial, four source buildings were selected based on their similar data distributions to the target building. These 4 source buildings were divided into two source domains for the TL task. Two Long Short-Term Memory (LSTM) models were trained on each source domain, and their

forecasting results for the target building were averaged after fine-tuning. This strategy resulted in a significant improvement in forecasting accuracy, enhancing predictions by 6.88 %–15.37 %.

In [22], the authors used measurement data of one source building from the open-source Building Genome Project [23] to support the energy prediction for two target buildings, also equipped with measurements through feature-based TL. All buildings involved are office buildings with comparable operational periods, ensuring that the datasets share similar distributions. This setup led to a notable improvement in model performance, with the accuracy increasing by over 15 % compared to a model that did not utilize TL.

Another study utilized the comprehensive dataset from the same Building Genome Project [23] to predict building energy consumption using TL integrated with a Multiple Layer Perceptron (MLP) model. This study analyzed a total of 8080 cases, selecting information-poor buildings as source buildings to enhance the forecasting accuracy for target buildings. The results demonstrated significant improvements in model performance attributable to the application of TL [24].

In [25], a TL framework was proposed for forecasting 12-h-ahead building energy consumption, leveraging pre-trained models for weight initialization in a combined convolutional neural network and LSTM architecture. All buildings involved were office buildings with measurement data. The results demonstrated an average improvement in model performance, as measured by the CVRMSE, of 19 %, with the best performance achieving a 23.64 % enhancement.

Three office buildings with the same operational schedule, located in close proximity across two cities, were selected for forecasting building electric energy consumption using TL [26]. TL was performed by fine-tuning model parameters, with certain neural network layers frozen. This paper utilized a LSTM-based model and achieved an average CVRMSE improvement of 31.18 %. The results concluded that pretraining a model on different buildings in the source domain can significantly enhance the accuracy of building energy predictions to a notable degree.

TL with a MLP model was utilized to predict energy consumption for four school buildings with measurement data [27]. The four schools had similar seasonal energy consumption patterns but varied in size and location across eastern Canada. Each school's dataset contained 17 features, which were used as inputs to a MLP model for energy forecasting. Integrating data from the source schools improved prediction accuracy by up to 11.2 %, compared to using only the target school's data for training.

Three TL strategies were used in Ref. [28] to improve the energy prediction performance across two building categories: office and institutional buildings. In each category, five buildings were designated as source buildings, while one building was selected as the target building. The datasets of measurements for all buildings were sourced from the Open-source Building Genome Project [23], and a LSTM model was used for prediction. Results demonstrated that using TL improved prediction accuracy by 75 % compared to models trained without TL.

The study by Yuan et al. [29] focused on predicting the one-day-ahead energy demand of a large-scale shopping mall, using measurement data from three other malls located in different climate zones as the source domain. The parameters from models trained on the source buildings were transferred to the target building model as the initialized parameters. The results indicated that, with careful selection of source domain data, TL significantly improved the performance of energy demand prediction compared to models without TL.

Another study [30] used TL to forecast the building energy use by integrating a multi-source buildings method. The similarity between the target building and multiple source buildings, extracted from the Open-source Building Genome Project [23], was assessed to guide the selection of source-building data for the TL task. The results verified the effectiveness of the multi-source TL method, improving model performance by 15 % compared to models that were derived solely from the target building data.

Table 1
Summary of the literature review of transfer learning for building energy forecasting.

Ref.	Building Type		Time step	Time lag	Forecast horizon	Method	Accuracy
	Source building	Target building					
[21]	Commercial, Office, and Industrial	Office	1 h	24 h	1 h	LSTM	CVRMSE of 25.37 % for the residential building; CVRMSE of 16.05 % for the office building.
[22]	Office	Office	1 h	24 h	1 h	LSTM	CVRMSE of 8.11 %
[24]	Institutional	Institutional	1 h	24 h	1 h	MLP	MAPE of about 10 % on average
[25]	Office	Office	1 h	2000 h	24 h	CNN and LSTM	CVRMSE of 23.64 % on average
[26]	Office	Office	1 day	1 month	1 day	CNN and LSTM	CVRMSE of 17.43 % on average
[27]	School	School	1 day	NA	1 month	MLP	MAPE of 21.49 % on average
[28]	Office and institutional	Office and institutional	1 h	24 h	1 h	LSTM	CVRMSE of 22.40 % on average for the office building; CVRMSE of 19.00 % on average for the institutional building;
[29]	Shopping mall	Shopping mall	1 h	NA	24 h	ARNN and LSTM	CVRMSE of 22.40 %
[30]	Office	Office	1 h	24 h	1 h	LSTM and DANN	CVRMSE of 11.47 %

Note: NA means Not Available, MAPE is the Mean Absolute Percentage Error, CNN is the Convolutional Neural Network, ARNN is the Auto-regression Neural Network, and DANN is the domain adversarial neural network.

2.2. Research gaps and contribution

Table 1 summarizes the reviewed studies [21,22,24–30] that primarily focused on transferring knowledge from source buildings with measurements to target buildings with also measurements, aiming to improve energy consumption forecasting. The literature review was conducted based on several criteria, such as publications within the past five years and a focus on building energy forecasting using neural network models combined with transfer learning techniques. Though significant efforts have been made to be comprehensive, it is acknowledged that not all relevant publications on this topic may have been included. Few studies explore the use of synthetic data from source buildings in TL tasks with real measurement data from target buildings. This limited application might be due to the difficulty in addressing gaps between the source and the target building, the discrepancies of input features between the two buildings (e.g., one necessary feature may not exist in another building), and differences in data quality between the synthetic and measured data. The literature review identifies only one paper [10] that used the DesignBuilder program [31] to simulate a building and used the simulated results as auxiliary training data to enhance energy forecasting for the same building with insufficient measurement data. This approach can be regarded as an application of data augmentation. Moreover, calibrating a DesignBuilder case or similar building performance simulation tools is a time-intensive and resource-demanding process.

Although synthetic data from Building Performance Simulation (BPS) tools is getting more easily accessible, its application to TL tasks has not been thoroughly investigated. This paper bridges the gap by applying TL from the source building with synthetic data to the target building with measurement data.

Three target variables are investigated in this paper: (1) the electrical power of the whole building (P_{BLDG}), (2) the electrical power of the whole HVAC system (P_{HVAC}), and (3) the average indoor air temperature (T_{ia}). P_{BLDG} and P_{HVAC} typically have large values and significant variance, whereas T_{ia} generally has smaller but more stable values. Estimating the three variables using a single model poses a challenge not only due to their distinct scales and variances, but also due to differences in modelling focus across disciplines. While normalization can mitigate scale-related issues in machine learning (ML) models, this challenge is more pronounced for non-ML models. Moreover, different perspectives lead to distinct research objectives: building engineers often focus on forecasting indoor air temperature to support thermal comfort and control strategies, whereas electrical engineers are more concerned with predicting total building energy consumption for load management and optimization. For instance, a study focusing on optimizing HVAC control using a transfer learning strategy considered only indoor air condition forecasting as the primary objective [32], while another study [21] proposed a multi-source-model approach for forecasting the building load, without including indoor air temperature as a model output. This paper addresses this challenge by developing a unified model capable of accurately forecasting all three variables. Furthermore, the proposed models are designed for direct integration into BAS for control purposes, extending beyond the prediction or forecasting capabilities emphasized in most studies. Combining the proposed models with advanced MPC strategies is expected to deliver benefits such as improved energy efficiency and reduced peak power demand [33]. While MPC has demonstrated strong potential in real-world building applications, its effectiveness depends heavily on the accuracy and robustness of the underlying predictive models. The work presented in this paper contributes to enhancing predictive accuracy, thereby supporting more effective implementation of MPC.

3. Methodology

3.1. Transfer learning strategies

Fig. 1 illustrates, at a high level, the TL concept adopted in this work. The working mechanism of this TL strategy is to transfer knowledge between the source building and the target building.

(a) From a source building that has synthetic data to pre-train neural network models,

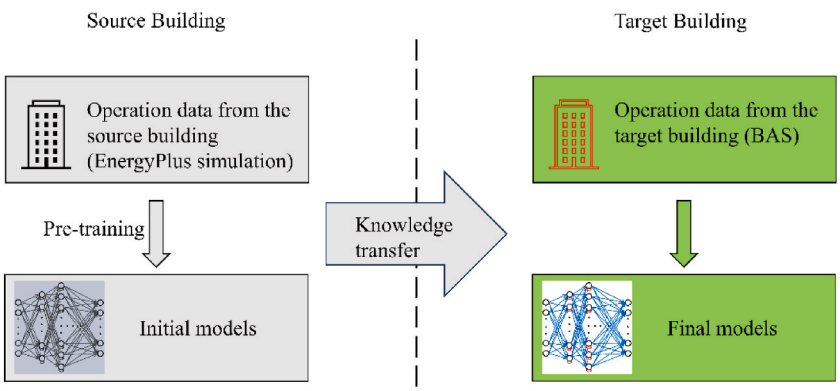

Fig. 1. Schematic of the studied transfer learning concept.

(b) to a target building that has measurement data to re-train (fine-tune) the neural network models and to validate the updated models.

Specifically, the TL strategy transfers the model parameters derived from the source building to serve as the initial model parameters for the target building. After fine-tuning with the target building data, these models are established and used to forecast the three variables of the target building.

This paper evaluates two TL strategies applied to the case study.

1. **DirectTL**: The neural network models are pre-trained and tested with the source building dataset, and then they are used directly with the dataset of the target building without changing the initial model weights. The performance of the neural network models is evaluated with the testing dataset of the target building.
2. **RefinedTL**: The neural network models are first pre-trained and tested with the dataset of source building, and then updated using the dataset of target building. The weights of all neural network model layers are fine-tuned with the training dataset of the target building. The updated neural network models are evaluated with the testing dataset of the target building.

3.2. Forecasting models

3.2.1. Available variables of datasets

The proposed neural network models are developed based on the available data of the source building (synthetic data from EnergyPlus simulation) and the target building (measurement data from BAS). These variables are used to formulate the neural network models and are selected as the model inputs and outputs.

The variables used in this paper for developing models are listed in Table 2 and are available in both the synthetic dataset of the source building and the BAS measurements of the target building. These variables are classified into four groups, including (1) building outdoor conditions, (2) building indoor environment, (3) building power demand, and (4) extracted variables. The first three groups (building outdoor conditions, building indoor environment, and building power demand) are directly from the datasets of the source building and target building, and are referred to as primary variables; while extracted variables are created as proxies and derived from the primary variables.

1. Solar air temperature T_{sa} is defined as the outdoor air temperature which, in the absence of solar radiation, would give the same temperature distribution and rate of heat transfer through a wall (or roof) as exists due to the combined effects of the actual outdoor temperature distribution plus the incidence solar radiation [34]. Thus, it is a crucial factor influencing building cooling load, affecting both P_{HVAC} and P_{BLDG}. It is described by Equation (1).

$$T_{sa} = T_o + \frac{A_{ss} SR}{h_o} \tag{1}$$

Where A_{ss} is the solar surface absorptivity, h_o is the coefficient of heat transfer by long-wave radiation and convection on an exterior surface, and SR is the solar radiation. In this study, $A_{ss} = 0.63$ and $h_o = 18$ W/m^{2}K [35].

2. Another extracted variable $\Delta T_{o,i}$ represents the temperature difference between indoor air and outdoor air, as defined in Equation (2). This variable represents the driving force of heat transfer between the indoor and outdoor environments, leading to the heat entering the building and consequently raising the HVAC system output. Therefore, $\Delta T_{o,i}$ directly affects building cooling load and, in turn, influences both P_{HVAC} and P_{BLDG} [36].

$$\Delta T_{o,i}^{t} = T_{oa}^{t} - T_{ia}^{t} \tag{2}$$

3. The time of the dataset is modified into a sine-wave-based variable as a regressor input, which is described in Equation (3) and Fig. 2. *Occ* here represents the modified time as a model input variable.

Table 2
Variables used for the modelling and transfer learning tasks.

	Variable groups	Variables	Descriptions	Unit
Primary variables	Building outdoor conditions	T_{oa}	Outdoor air temperature	°C
		RH_{oa}	Relative humidity of outdoor air	%
		SR	Solar radiation	W/m^2
	Building indoor environment	T_{ia}	Indoor air temperature	°C
		$T_{ia,sp}$	Setpoint temperature of indoor air	°C
	Energy consumption	P_{HVAC}	Electric power of the whole HVAC system (HVAC power)	kW
		P_{BLDG}	Electric power of the whole building (Total building power)	kW
Extracted variables		Occ	Modified time as a proxy for occupancy	1
		$\Delta T_{o,i}$	Temperature difference between outdoor and indoor air	°C
		T_{sa}	Solar air temperature	°C

$$Occ = \text{abs}(\sin(Time)) \tag{3}$$

The absolute operation is used in Equation (3) to avoid negative values with a period of 24 h. Thus, daily time is modified into the variable *Occ*. During business hours (from 8:00 a.m. to 5:00 p.m.), buildings are typically occupied, with occupancy levels often peaking gradually around noon. Fig. 2 begins at 7:00 a.m., reflecting the typical early start of the HVAC system to precondition the indoor environment and ensure the temperature reaches the desired comfort level before occupants arrive. Therefore, the values of *Occ* during business hours can be represented by sine wave values, while those during unoccupied hours (before 7:00 a.m. and after 5:00 p.m. daily) are set to zero, as depicted in Fig. 2. As for the daily patterns of building energy consumption, it is typically case-dependent and time-dependent. For instance, P_{BLDG} in an office building usually rises at 9:00 a.m. and peaks around 4:00 p.m. [37]. Thus, it is expected that this extracted variable *Occ* can represent, to some extent, both building occupancy and energy-related patterns.

The variables introduced in this section undergo a selection process based on Pearson coefficient analysis to identify the most relevant input variables for enhancing model performance. Further details are discussed in Section 5.1.

3.2.2. Neural network models

This paper uses three neural network models to forecast the target variables. Neural network models are computational frameworks composed of interconnected neurons arranged in layers. Each neuron computes a weighted sum of its inputs and applies an activation function to introduce non-linearity. This structure enables neural networks to model complex, non-linear relationships between inputs and outputs. They can automatically learn feature representations from raw data, making them suitable for a diverse range of applications [38].

The three neural network models utilized in this paper are multiple-layer perceptron (MLP), Long-and-short term memory (LSTM), and gated recurrent unit (GRU). These three models are selected due to their high performance as reported in Refs. [39,40]. MLP works in a feed-forward way, and it contains one or more hidden layers, where each layer has one or more neurons [41]. The MLP model can achieve the best performance with (i) two hidden layers, (ii) the number of neurons of the first hidden layer equalling $2 \times (2 \times n+1)$, where n is the number of neurons of the input layer, and (iii) the number of neurons of the second hidden layer equalling $(2 \times n +1)$ [42]. The governing function and the definition of the hidden layer of a MLP are presented in Ref. [17]. The LSTM model and GRU model are two recurrent neural network models excelling in time-series tasks, and their governing equations are presented in Ref. [43]. For the LSTM model focusing on the regression task of forecasting HVAC energy consumption, its structure with (i) one single LSTM lay with 4 neurons and (ii) hyperbolic tangent sigmoid activation function (Tanh, Equation (4)) performs best [44]; therefore, this paper utilized this architecture to build the LSTM model. The optimum architecture of the GRU model used in this paper consists of one single GRU layer with 5 neurons and tanh activation function (Equation (4)) referring to Ref. [45]. The optimum structure of each neural network model is summarized in Table 3.

$$tanh\,(x) = \frac{exp(x) - exp(-x)}{exp(x) + exp(-x)} \tag{4}$$

3.2.3. Modelling

This paper builds forecasting models using neural network techniques, as illustrated in Fig. 3. The objective is to evaluate the 12-h-ahead forecasting of the electric power of the whole HVAC system (P_{HVAC}) and the whole building (P_{BLDG}) during the cooling season. Additionally, to ensure the building's indoor thermal comfort, the average indoor air temperature T_{ia} is also considered as a model

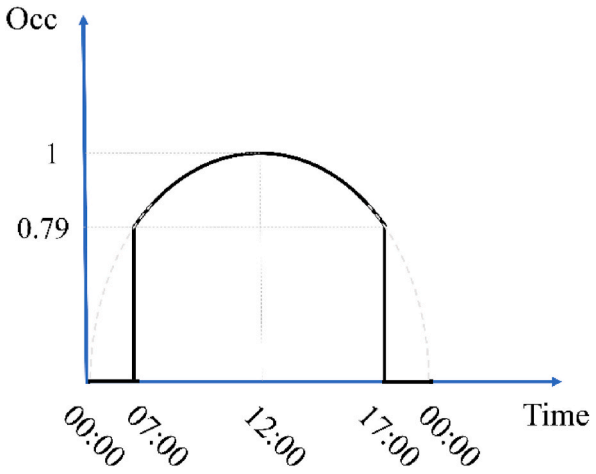

Fig. 2. Daily pattern of the extracted variable of *Occ*.

Table 3
Optimum architecture of neural network models.

Neural network models	Description of the optimum structure
MLP	Assume n neurons on the input layer. i. $2 \times (2 \times n+1)$ neurons on the first hidden layer. ii. $2 \times n+1$ neurons on the second hidden layer.
LSTM	i. one single LSTM layer with 4 neurons. ii. activation function: tanh.
GRU	i. one single GRU layer with 5 neurons ii. activation function: tanh.

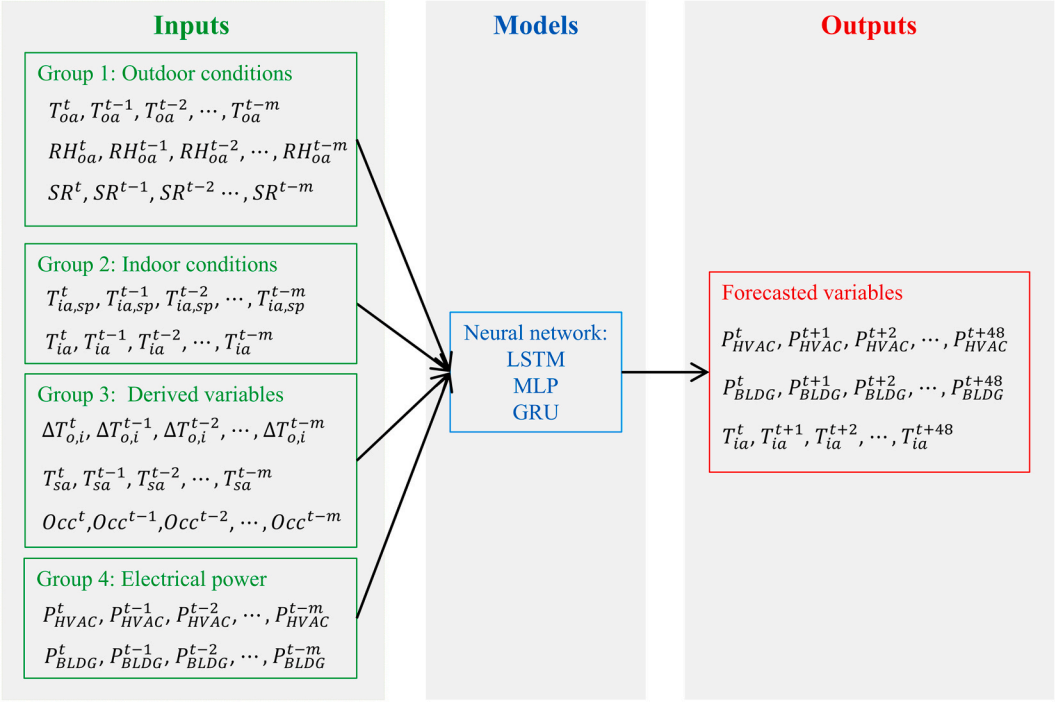

Fig. 3. The architectures of three neural network models used in the transfer learning with a 12-h forecasting horizon and a 15-min interval.

output. Energy of the whole HVAC system includes the electrical energy of the primary cooling plant (chillers, pumps, cooling towers, etc.) and secondary cooling systems (air-handling units, fan coils, pumps, fans, etc.). The total building power (P_{BLDG}) represents the electrical energy of the whole building.

Two types of strategies are commonly used for solving multi-step-ahead forecasting problems: (i) iterative strategy and (ii) direct strategy [46,47]. The iterative strategy employs the same forecasting model to predict the target variable one time step ahead repeatedly and sequentially until the desired forecast horizon is reached. As a result, forecasting errors accumulate iteratively and potentially diminish model performance over longer horizons. In contrast, the direct strategy trains a single model to forecast multiple time steps ahead simultaneously. By avoiding error propagation, this approach is typically more accurate and efficient for multi-step-ahead forecasting [44]. This paper uses the direct strategy to forecast 12-h ahead building energy consumption. The outputs of forecasting models (energy consumption of the whole HVAC system, energy consumption of the whole building, and T_{ia}) extend from the current time step t to the desired forecast horizon (12 h), for each target variable. As the time step of the case study in this work is 15 min, the output variables expand from t to $t+48$ (12 h) in Fig. 3. The historical measurements are represented by m in Fig. 3, indicating the number of time lags a model considers.

The model inputs are classified into four groups as depicted in Fig. 3, they are **historical measurements of** (1) building outdoor conditions, (2) building indoor environment, (3) extracted variables, and (4) energy consumption.

Input Group 1 is building outdoor conditions, including three variables: T_{oa}, RH_{oa}, and SR. They represent the historical meteorological effects on the space cooling demand and operation of the HVAC system (e.g., building operators may take T_{oa} as the indicator to turn on/off the HVAC system).

Input Group 2 consists of $T_{ia,sp}$ and T_{ia}, and they represent the impacts of historical indoor conditions and control actions on future building energy and indoor environment.

Input Group 3 comprises three extracted variables $\Delta T_{o,i}$, T_{sa}, and *Occ*. They represent the historical combined effects of various factors on the model targets. For instance, the historical values of $\Delta T_{o,i}$ (see Equation (1)) reflect the integrated impacts of factors such as heat transfer driving force, building envelope heat capacity, etc., on building energy and indoor thermal comfort.
Input Group 4 consists of the historical measurements of P_{HVAC} and P_{BLDG}. Since this paper evaluates the future behaviour of these two variables, their historical measurements are expected to significantly influence their future values.

As the scale of model input and output variables in source building and target building differs a lot, the min-max normalization method is used in this paper to adjust variable values to a range of 0 and 1. The equation for the min-max normalization of each variable is presented in Equation (5):

$$X_{i,norm} = \frac{X_i - X_{i,min}}{X_{i,max} - X_{i,min}} \tag{5}$$

Where X_i is a variable of source and target buildings, $X_{i,norm}$ is the normalized value of X_i, X_{min} is the minimum value, and X_{max} is the maximum value.

These neural network models are developed using Python (version 3.9.12) [48]with open-source libraries like Tensorflow (version 2.10.0) [49].

3.3. Performance evaluation

Two performance metrics, Root-Mean Square Error (RMSE) and Coefficient of Variation of Root-Mean Square Error (CVRMSE), are defined in Equations (6) and (7), which are used to evaluate the model performance. Here, y_i is the measured value, $\widehat{y_i}$ is the forecasted value, $\bar{y}$ is the mean value of measurements, and n is the sample size.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n}} \tag{6}$$

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}{n}}}{\bar{y}} \times 100\% \tag{7}$$

To evaluate the relative performance of the two TL strategies, other neural network models are developed by using the so-called self-learning (**SelfL**) strategy. Specifically, three neural network models (LSTM, MLP, and GRU) are trained and tested only with the target building dataset. The SelfL strategy is to assess what performance a neural network model can achieve using only the limited dataset of the target building. In this case, it helps to evaluate if the TL strategies are necessary by incorporating datasets from the source building. Thus, models developed with the SelfL strategy serve as the baseline models to be compared with other neural network models derived from the DirectTL and RefinedTL strategies.

To describe the model performance improvement, we adopt the Ratio of Relative Performance Improvement (RRPI) as defined in Equation (8). $RMSE_{SelfL}$ is the root mean squared error over the testing dataset of the target building using the SelfL strategy, while $RMSE_{TL}$ is the root mean squared error over the testing dataset of the target building using the TL strategies. Thus, RRPI serves as an

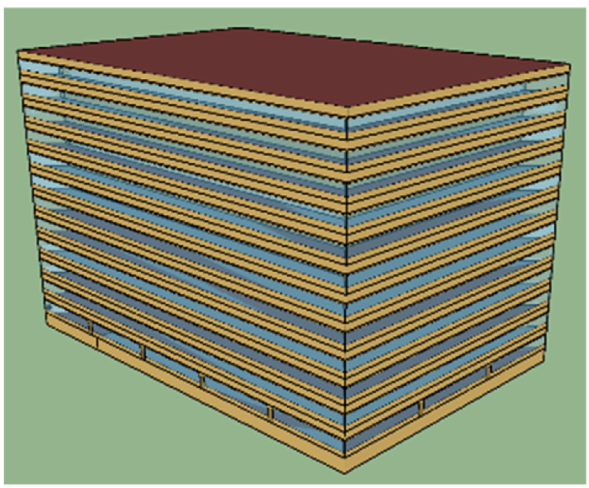

Fig. 4. The source building from an open-source building prototype.

indicator to evaluate the effect of knowledge transfer on the model enhancements. Larger positive values of RRPI indicate a more significant effect of TL.

$$RRPI = \frac{RMSE_{SelfL} - RMSE_{TL}}{RMSE_{SelfL}} \times 100\% \quad (8)$$

Table A1 outlines the pseudocode for the transfer learning-based forecasting workflow used in this paper.

4. Case study

4.1. Datasets of source and target buildings

The selected source building data encompasses three months (June, July and August) at 15-min time intervals, while the target building data encompasses two months (July and August) of 2020 at 15-min time intervals. Weekends and holidays in both the source building data and the target building data are removed for model training and testing.

The source building (see Fig. 4) has 12 floors with a total floor area of 46,320 m^2. The model is the large office archetype from the U. S. Department of Energy (DOE) prototype buildings [14]. The vintage of the selected model is based on the 2022 version of the ASHRAE Standard 90.1 [50], while the location is Rochester, Minnesota, U.S. The HVAC system in the source building consists primarily of variable air volume boxes with reheat and two chillers connected to two variable-speed cooling towers, serving 23 conditioned zones. The simulation results from EnergyPlus, using weather data from the typical meteorological year, serve as the dataset for the source building. This approach is effort-efficient to provide essential data to support forecasting for a target building. The setpoint temperatures and indoor air temperatures from all thermal zones in the source building are averaged to represent the whole-building setpoint and indoor air temperature. All model input data are then normalized before being fed into the model, using the method described in Section 3.2.

The target building is a commercial building with a floor area of 36,000 m^2 and 12 floors, located in downtown Montreal, Canada. It was first built in 1913 with a new extension built in 2016. The cooling plant of the target building consists of three chillers (one centrifugal chiller and two screw chillers) connected to four cooling towers. The airside system comprises two fresh air pre-treatment units as the primary Air Handling Unit (AHU) system and 22 cooling coils as the secondary AHU system. The BAS data used in this work were from 2020. A detailed description of the target building and its HVAC systems is presented in Ref. [51]. The weather data, including solar radiation for the target building, comes from publicly available measurements [52]. The target building data is first pre-processed by filling up recording errors using linear interpolation with the Pandas library [53].

Note that the source building is intentionally selected to differ from the target building, except for the building type. The significant differences between the source building and the target building provide an opportunity to evaluate the transferability of neural network models when substantial differences exist between the source and target domains.

4.2. Model development

As introduced above, the data over both the source and target buildings covers only two or three months. To ensure the neural network models are exposed to a comprehensive range of conditions, the models are first pre-trained using the entire dataset of the source building. By doing so, the pre-trained model gains exposure to August conditions, which can contribute valuable knowledge when transferred to the target building, potentially enhancing forecasting performance. The August data from the target building is reserved for testing the models, ensuring an unbiased evaluation of their forecasting performance. To prevent over-fitting, the neural network models are only trained for a maximum of 20 epochs on source building. Given the discrepancy between the source and target buildings, pre-training these neural network models with a larger loss function may help capture the underlying energy patterns of the building while allowing the models to be more easily adjusted with the target building data.

For the target building, the dataset is divided into a set of training, validation, and testing subsets, as listed in Table 4. Periods 1

Table 4
Period of training dataset, validation dataset, and testing dataset over the source building dataset and the target building dataset.

Datasets	Data type	Dataset division		Duration	Number of days	Dataset size
Source building	Synthetic	/	/	06/01–08/31	64	6144
Target building	Measured	Period 1	Training dataset	07/01–07/10	7	768
			Validation dataset	07/11-07/18	5	488
			Testing dataset	07/19-08/31	31	2976
		Period 2	Training dataset	07/01–07/13	8	864
			Validation dataset	07/14-07/25	9	864
			Testing dataset	07/26-08/31	26	2496
		Period 3	Training dataset	07/01–07/14	9	960
			Validation dataset	07/15-07/25	8	768
			Testing dataset	07/26-08/31	26	2496
		Period 4	Training dataset	07/01–7/15	10	1056
			Validation dataset	07/16-07/25	7	672
			Testing dataset	07/26-08/31	26	2496

through 4 progressively increase the training dataset by one day each. For example, Period 1 has a sample size of 768 (covering 8 days), while Period 2 includes 9 days of data. This method of data division is designed to simulate scenarios where a building is newly constructed or recently retrofitted, resulting in very limited data availability.

5. Results and discussion

5.1. Analysis of source and target buildings

The Pearson coefficient [54] between the three target variables and a group of regressor variables is calculated. Table 5 presents the absolute values of the Pearson coefficient, indicating the correlation levels between the target variables and the regressor variables. Overall, T_{ia} has relatively weak correlations with regressor variables compared to P_{HVAC} and P_{BLDG}. This could be due to T_{ia} being more stable with less variance. Pearson coefficients with absolute values greater than 0.5 are typically associated with accurate time-series modelling [55], and this threshold is used in this paper to select regressor variables. Therefore, all regressor variables in Table 5 are chosen as model inputs because their Pearson coefficient with at least one of the three target variables exceeds 0.57. For example, while the Pearson coefficients of RH_{oa} with T_{ia} and P_{BLDG} are 0.12 and 0.48, respectively, its correlation with P_{HVAC} is 0.57 (greater than 0.5), making RH_{oa} a selected input variable for the model.

Table 6 presents the statistical indicators, including the average values and standard deviations, of all primary and extracted variables across the entire datasets of source and target buildings. Overall, variables of T_{oa}, RH_{oa}, $T_{ia,sp}$, T_{sa}, and *Occ* show very close similarities between source and target buildings in terms of both average values and standard deviations. The mean values of $\Delta T_{o,i}$ over source building show a negative value, indicating that during most times, particularly at night, the outdoor air temperature in summer is lower than T_{ia} in Rochester. Although the mean value of $\Delta T_{o,i}$ over the target building is slightly positive at 0.19 °C, this is likely because the target building data does not cover June, a month when T_{oa} in Montreal are relatively cool compared with July and August, especially at night. For the target variables, T_{ia} and P_{HVAC} show similar values across the source building and target building, indicating similar patterns in terms of indoor thermal conditions and electric power consumption of the HVAC system. However, P_{BLDG} shows a significant difference with P_{BLDG} in the source building being more than twice that in the target building. In particular, the source building simulated in EnergyPlus outputs a higher share of non-HVAC power (68 % vs. 17 %) due to the use of default power densities and fixed operation schedules for non-HVAC equipment such as lighting, appliances, and plug loads. On the other hand, the target building, a real building, is subject to human interventions from building managers and occupants, particularly their behaviour on the use of lights, appliances and plug loads.

Fig. 5 provides an example of five executive days to illustrate the patterns of P_{HVAC} and P_{BLDG}, along with the outdoor conditions of T_{oa} and solar radiation of the source and target buildings. The data presents some similar trends of the source building and the target building, with both P_{HVAC} and P_{BLDG} peaking around noon and reaching their lower level at midnight. However, some differences are also noticed: (1) the higher-level P_{HVAC} and P_{BLDG} in the target building usually persist longer than those in the source building, and (2) P_{HVAC} in the target building tends to align more closely with P_{BLDG} compared to the source building. The variance of HVAC power and total building power is noticed in Fig. 5(a), especially during the night.

The outdoor air temperature T_{oa} and solar radiation show a significant impact on the HVAC power and total building power of the source and target buildings, as indicated in Fig. 5(a) and (b). For instance, statistical analysis indicates, on average, P_{BLDG} increases by 85 kW for the whole building and by 95 kW for P_{HVAC} per 5 °C of increase in outdoor air temperature over the training and validation dataset of the target building. This indicates the HVAC power is more sensitive to the outdoor air temperature compared with the total building power. This characteristic aligns well with the fact that the building typically integrates multiple systems that are not affected by outdoor conditions, such as the lighting system.

5.2. Results of transfer learning

5.2.1. Overall performance of transfer learning

Each of the three neural network models (LSTM, MLP, and GRU) is configured with the optimum structure (Table 3) and pre-trained using source building with the synthetic data generated from EnergyPlus simulations (Table 4). Then, those models are fine-tuned and tested on four periods of target building (Table 4) using different TL strategies. Specifically, this paper evaluates two TL strategies and

Table 5
Pearson correlation between target variables and regressor variables on source building data.

Regressor variables	Target variables		
	T_{ia}	P_{HVAC}	P_{BLDG}
T_{oa}	0.17	0.84	0.68
RH_{oa}	0.12	0.57	0.48
SR	0.34	0.68	0.66
$T_{ia,sp}$	0.72	0.62	0.64
T_{sa}	0.21	0.81	0.74
$\Delta T_{o,i}$	0.08	0.87	0.71
Occ	0.33	0.69	0.72

Table 6
Statistical summary of primary variables and extracted variables over the entire dataset of the source and target buildings.

Variables	Source building		Target building	
	Mean	Standard deviation	Mean	Standard deviation
T_{oa} (°C)	20.01	4.95	23.31	4.26
RH_{oa} (%)	75.25	15.95	70.80	20.39
SR (W/m^2)	314.87	345.95	240.54	293.17
T_{ia} (°C)	24.85	0.45	23.12	0.26
$T_{ia,sp}$ (°C)	25.37	1.01	24.11	1.35
T_{sa} (°C)	25.6	9.71	53.62	39.08
$\Delta T_{o,i}$ (°C)	−4.84	4.89	0.19	4.23
Occ (−)	0.42	0.46	0.40	0.46
P_{HVAC} (kW)	296.71	177.71	345.09	124.26
P_{BLDG} (kW)	938.65	322.33	418.92	117.84

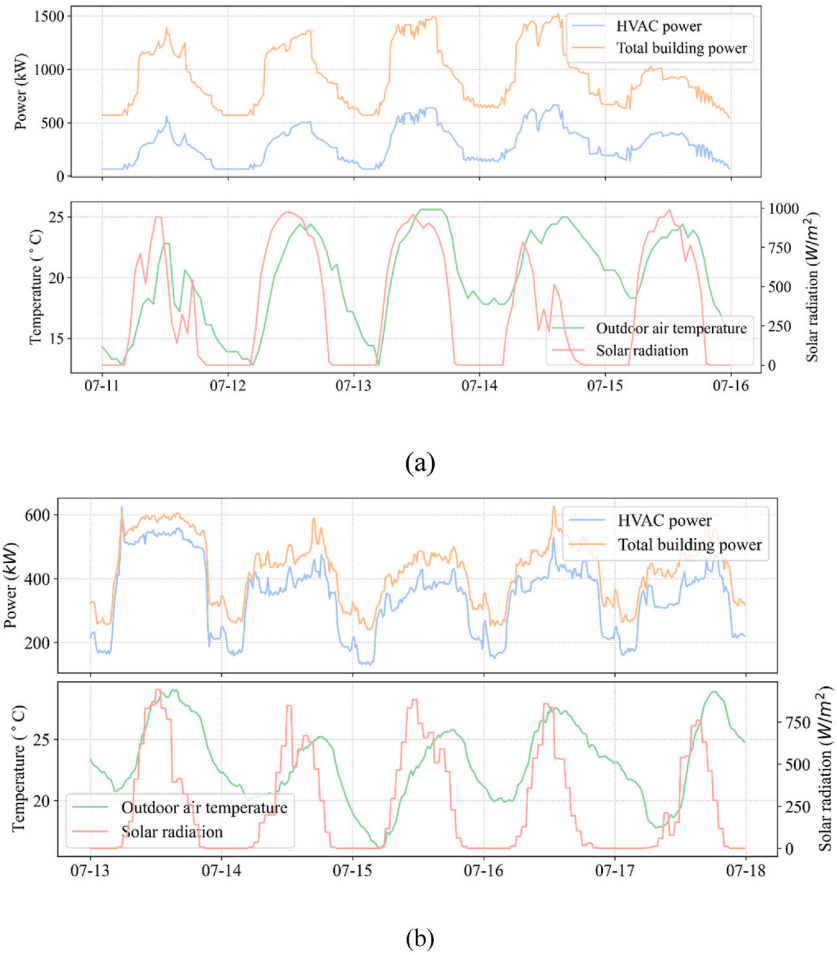

Fig. 5. Example of HVAC power and total building power in five consecutive days. (a) source building; (b) target building.

one SelfL strategy applied across three neural networks with four periods of target building, resulting in a total of thirty-six cases. The results of those cases are presented in Table 7. As the analysis in this paper finds time lags do not significantly affect the model performance, the results focus specifically on cases with a 24-h time lag. Overall, as more data becomes available for model re-training (Table 4), the model's performance tends to increase (see Table 7). Table 7 also indicates the CVRMSE for total building power in each case is lower compared to the corresponding CVRMSE for HVAC power. This difference is attributed to the target building's integration of additional systems, such as internal lighting loads, which maintain relatively steady power demands. As a result, the overall variation in total building power is reduced.

According to ASHRAE Guideline 14 [56], the calibration of a computer model of a single building use is considered acceptable if the

Table 7
Results of RMSE and CVRMSE for the three target variables over the testing dataset of the target building.

Datasets	Transfer learning strategies	LSTM			MLP			GRU		
		RMSE (°C)	CVRMSE (−)		RMSE (°C)	CVRMSE (−)		RMSE (°C)	CVRMSE (−)	
		T_{ia}	P_{HVAC}	P_{BLDG}	T_{ia}	P_{HVAC}	P_{BLDG}	T_{ia}	P_{HVAC}	P_{BLDG}
Period 1	SelfL	0.25	0.33	0.27	0.22	0.22	0.18	0.26	0.34	0.28
	DirectTL	0.37	0.54	0.29	0.33	0.33	0.19	0.37	0.47	0.24
	RefinedTL	0.22	0.22	0.17	0.20	0.20	0.16	0.24	0.23	0.18
Period 2	SelfL	0.27	0.39	0.32	0.21	0.22	0.17	0.26	0.35	0.29
	DirectTL	0.41	0.55	0.30	0.37	0.31	0.19	0.41	0.49	0.24
	RefinedTL	0.30	0.26	0.21	0.20	0.20	0.15	0.19	0.23	0.19
Period 3	SelfL	0.26	0.26	0.22	0.21	0.21	0.17	0.26	0.33	0.27
	DirectTL	0.41	0.55	0.30	0.37	0.31	0.19	0.41	0.49	0.24
	RefinedTL	0.23	0.21	0.16	0.20	0.19	0.15	0.18	0.22	0.17
Period 4	SelfL	0.24	0.23	0.19	0.20	0.20	0.16	0.25	0.30	0.25
	DirectTL	0.41	0.55	0.30	0.37	0.31	0.19	0.41	0.49	0.24
	RefinedTL	0.23	0.21	0.16	0.19	0.19	0.14	0.18	0.21	0.15

CVRMSE between the measurements and predictions of building energy is smaller than 30 % when using hourly data. Despite the use of a 15-min dataset, where greater variance is expected compared to hourly data, this paper achieves better results, with the CVRMSE of P_{BLDG} remaining below 21 % for all the investigated cases using the Refined TL strategy. The best performance is achieved with the MLP model, achieving a CVRMSE of 14 %. These findings demonstrate that the transfer learning strategy, supported by neural network models, is effective for accurately forecasting building energy.

RRPI is calculated with respect to SelfL and RefinedTL and presented in Table 8. The discussion of the comparison of SelfL and DirectTL is presented in Section 5.2.2. Positive RRPI values affirm the effectiveness of RefinedTL, though varying trends are observed with the accumulation of the training dataset on target building across the three neural network models. For instance, LSTM presents a declining trend in effectiveness for both P_{HVAC} and P_{BLDG}, which aligns with the findings of [24]. Whereas GRU shows an increasing trend for P_{BLDG}. The MLP model, in contrast, maintains relatively steady effectiveness for P_{HVAC} and P_{BLDG}. Those characteristics indicate that the knowledge extracted from the synthetic dataset (source building), despite differences in P_{HVAC} and P_{BLDG} patterns, enhances the forecasting performance for the target building with measurements. The TL strategy proves more effective with LSTM when the training dataset on the target building is small, while it performs better with GRU as the training dataset on the target building accumulates. The highest RRPI, noticed in Period 4, is 41 % using the GRU model, indicating the strongest improvement through TL. The average RRPI in Table 8 shows that TL integrating GRU results in the most significant improvement in the TL effect.

Fig. 6 presents the RMSE of T_{ia} using RefinedTL and SelfL strategies across three neural network models. Overall, the RefinedTL strategy performs slightly better than the SelfL strategy, regardless of the neural network model used. However, the RMSE differences between the two strategies are minor, as Fig. 6 indicates RMSE values of all investigated cases are small. These values even fall within the accuracy range of common temperature sensors used in most buildings [57], indicating that the forecasting of T_{ia} in this paper is accurate. The forecasted T_{ia} can serve as a key indicator for evaluating indoor thermal comfort when implementing actions (e.g., control strategies) aimed at enhancing building energy efficiency while having to consider thermal comfort.

5.2.2. *Performance of neural network models*

Fig. 7 displays the CVRMSE of HVAC power and total building power across the four periods of testing datasets of the target building resulting from the LSTM model. Overall, RefinedTL performs better than the DirectTL and SelfL strategies. The results of CVRMSE of the total building power indicate the performance of the SelfL strategy gradually increases and exceeds DirectTL as the training datasets accumulate. The best model performance is achieved with a CVRMSE of 21 % for P_{HVAC} and 16 % for P_{BLDG}, when 11 days' data (Period 4) is available for model re-training.

Fig. 8 illustrates the CVRMSE of HVAC power and total building power across the testing datasets of four periods of the target building using MLP models. Overall, the RefinedTL strategy outperforms the DirectTL strategy and SelfL strategy. In contrast, the DirectTL strategy shows the poorest performance for both HVAC power and total building power, indicating MLP models trained solely

Table 8
RRPI values in % resulting from RefinedTL over the target building.

Datasets	Neural network models					
	LSTM		MLP		GRU	
	P_{HVAC}	P_{BLDG}	P_{HVAC}	P_{BLDG}	P_{HVAC}	P_{BLDG}
Period 1	34	36	9	10	33	34
Period 2	32	35	9	11	34	35
Period 3	21	27	8	12	35	39
Period 4	8	18	6	12	29	41
Average	24	29	8	11	33	37

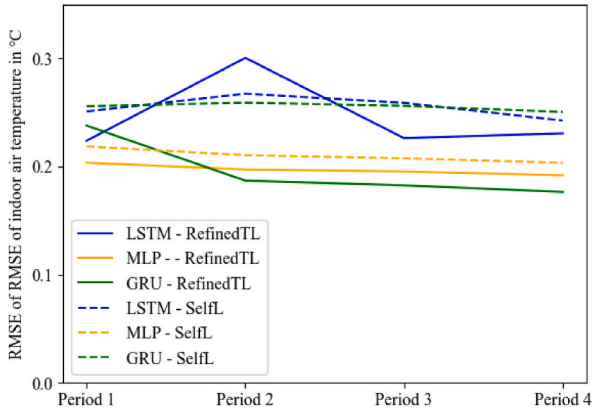

Fig. 6. RMSE of indoor air temperature over the testing dataset of four periods of the target building.

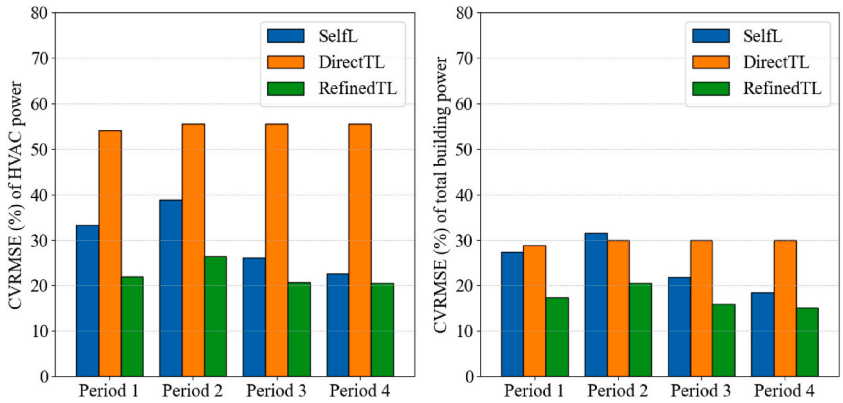

Fig. 7. CVRMSE of the HVAC power and total building power across the four periods of testing datasets of the target building using the LSTM model.

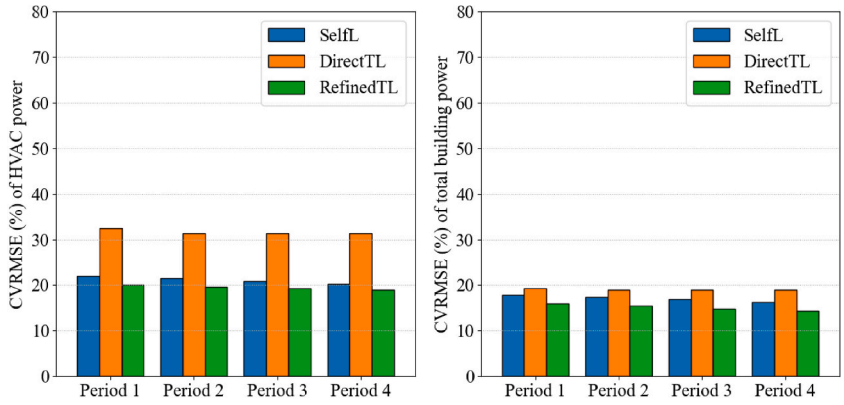

Fig. 8. CVRMSE of the HVAC power and total building power across the four periods of testing datasets of the target building using the MLP model.

on the data from the source building are inadequate for the accurate forecasting of the target building. A similar characteristic is noticed in the cases of the SelfL strategy across the four periods, which utilizes very limited datasets that include only the target building. The performance of the MLP model remains relatively stable as the amount of training data increases. The best performance is achieved in Period 4, with CVRMSE of 19 % and 14 % for HVAC power and total building power (Table 7) using the RefinedTL strategy.

Fig. 9 illustrates the CVRMSE of HVAC power and total building power across the four periods of testing datasets of the target building by using the GRU model. Overall, the RefinedTL strategy demonstrates the best performance, achieving CVRMSE of 21 % and

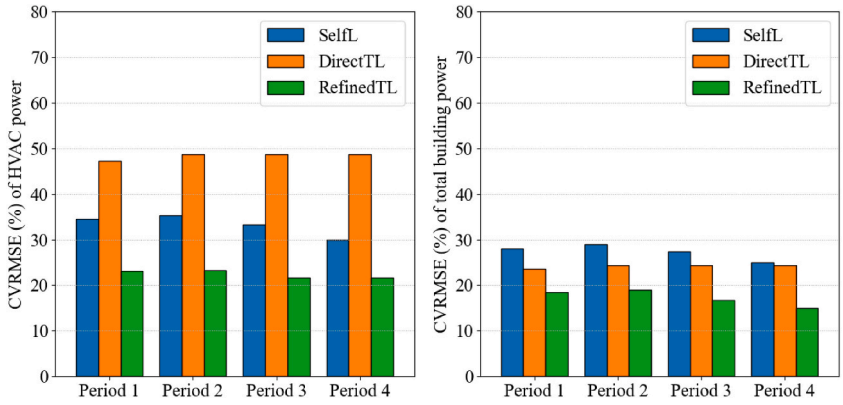

Fig. 9. CVRMSE of the HVAC power and total building power across the four periods of testing datasets of the target building using the GRU model.

15 % for HVAC power and total building power in the case of Period 4. For the cases with RefinedTL strategy, a significant model performance improvement with CVRMSE reaching 21 % is noticed for HVAC power when 11 days' data (Period 4) is used for model re-training. In contrast, the CVRMSE remains above 22 % for the other three cases of Periods 1–3, despite they have 8–10 days of data available for the model re-training. The addition of just one extra day of data makes a notable difference in model performance.

5.2.3. Temporal performance

Table 9 presents the computation times of the LSTM, MLP, and GRU models with respect to Period 4 of the target building as an instance. Overall, the RefinedTL strategy requires the most computation time compared to the DirectTL and SelfL strategies for each model. However, considering the performance improvements achieved with the RefinedTL strategy, the differences in computation times of the three strategies (DirectTL, SelfL, and RefinedTL) are not significant for preventing the use of the RefinedTL strategy instead of the DirectTL and SelfL strategies in practice.

Fig. 10 compares the measured and forecasted values of three target variables (T_{ia}, P_{HVAC}, and P_{BLDG}), resulting from the RefinedTL strategy using a LSTM model for an example week from 2020/07/27 to 2020/07/31, within the testing dataset of Period 4. The learning curve resulting from this case is presented in the Appendix (Figure A1). The large variance of HVAC power and total building power occurs in the morning when the HVAC system operates from the low-load level to the high-load level.

A visual inspection of Fig. 10 indicates the LSTM model supported by TL can forecast well the overall trends of total building power and HVAC power 12-h in advance, maintaining a reasonable accuracy. The forecasting of indoor air temperature is accurate with a RMSE of 0.23 °C (Table 7). However, some deviations are observed during periods of high and low power demand for both total building power and HVAC power. Table 7 indicates the overall accuracy of LSTM models as measured by CVRMSE are 16 % and 21 % for total building power and HVAC power, respectively, across the whole testing dataset of Period 4. These variations in accuracy can be attributed to several factors, including unpredictable human behaviour, occasional disturbances or malfunctions in HVAC operations, and the complex interactions between different building systems such as HVAC, lighting, and appliances.

Fig. 11 displays the comparison of measured and forecasted values of three target variables (T_{ia}, P_{HVAC}, and P_{BLDG}) using a MLP model with the RefinedTL strategy, covering an example week from July 27 to July 31, 2020, within the testing dataset of Period 4. The learning curve associated with this case is detailed in the Appendix (Figure A2). Overall, the forecasted values closely align with the measured values for T_{ia}, P_{HVAC}, and P_{BLDG}. While some deviations are noticed for HVAC power, the CVRMSE for the testing dataset of Period 4 is a low 19 % (Table 7). The forecasted indoor air temperature also tracks well with measured values, achieving a RMSE of only 0.19 °C (Table 7).

Fig. 12 compares the measured and forecasted values of three target variables (T_{ia}, P_{HVAC}, and P_{BLDG}), which are obtained using the RefinedTL strategy with a GRU model for an example week from July 27 to July 31, 2020, within the testing dataset of Period 4. The learning curve resulting from this case is presented in the Appendix (Figure A3). Fig. 12 indicates both the measured total building power and HVAC power show greater variability compared to the forecasted values. While the forecasted total building power and HVAC power generally follow the measured trends, some deviations are observed, particularly during peak hours in the daytime. The best performance is noticed in the case of Period 4 with CVRMSE of 21 % and 15 % for P_{HVAC} and P_{BLDG}, respectively. In this case, the RMSE of T_{ia} is 0.18 °C.

The best performance observed in our case study was a CVRMSE of 14 % for P_{BLDG} using the MLP model during Period 4. In comparison, another study employed transfer learning to forecast 1-h-ahead energy consumption for five office buildings, using measurement data from both source and target buildings. That study reported an average CVRMSE of 22.40 % [28], which is slightly less accurate than the results achieved in our case.

6. Conclusion

When applying data-driven models to building-related predictions or forecasting, a major challenge lies in the lack of sufficient

Table 9
Computation time for each model for Period 4.

Models	DirectTL	SelfL	RefinedTL
LSTM	15.38s	4.38s	23.83s
MLP	21.97s	7.63s	41.01s
GRU	17.71s	4.29s	45.09s

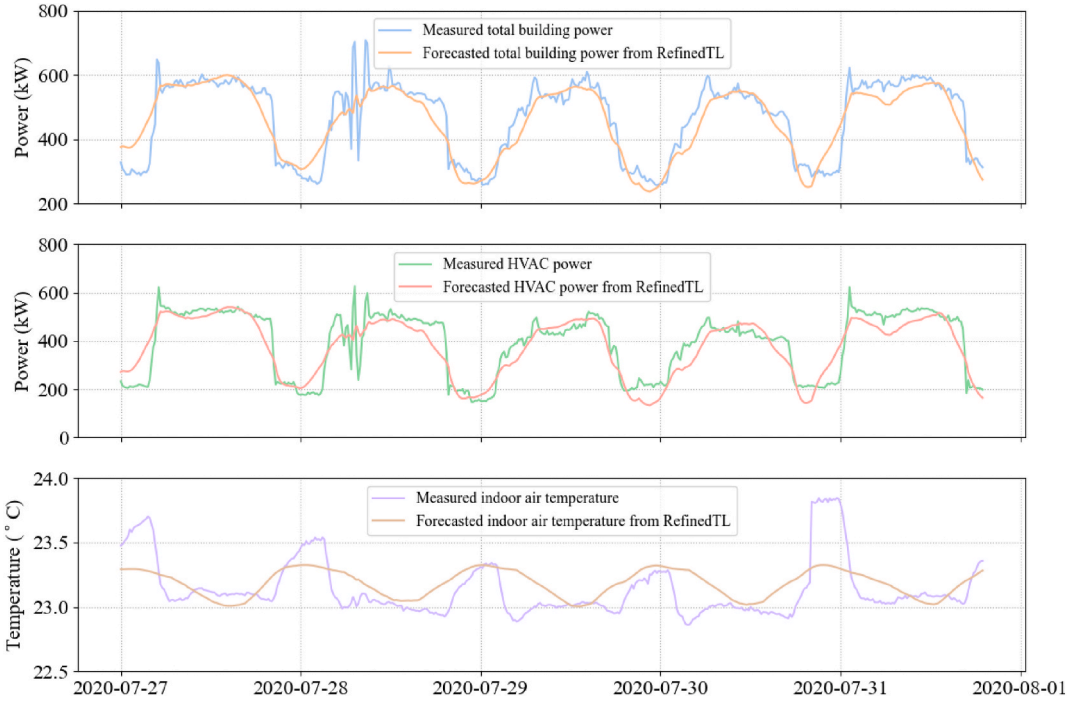

Fig. 10. Comparison of measured and forecasted total building power, HVAC power and indoor air temperature, resulting from the LSTM model using RefinedTL strategy for an example week in the testing dataset of Period 4 of the target building.

operation data, particularly for new or retrofitted buildings. This is especially true for models based on machine learning or artificial intelligence methods, which typically require large volumes of training and validation data to learn and generalize effectively. As a result, the absence of rich datasets makes it difficult to directly apply data-driven methods, e.g., for the controller model within MPC.

To address this challenge, this study investigated the application of transfer learning techniques to enhance the forecasting performance of black-box models under conditions of limited data availability. Transfer learning allows models to leverage knowledge gained from one domain or dataset to improve performance in another, making it particularly useful in cases where data from the target environment is scarce.

In this research, we utilized an effort-efficient approach to develop the source building and generate synthetic data using the EnergyPlus program to pre-train neural network models. These pre-trained models were then fine-tuned using a limited set of real measurements from an actual building. The real and synthetic buildings, though sharing the same architectural archetype, are located in different cities, each subject to unique weather and environmental conditions.

This study explored two distinct transfer learning strategies and evaluated their effectiveness using three different neural network models. By comparing the results, we aimed to determine how well transfer learning techniques can improve the accuracy of the forecasting models. The simulation results revealed that while the performance of the models varies depending on the transfer learning approach and the specific neural network architecture employed, the incorporation of synthetic data in the pre-training phase consistently leads to significant improvements in forecasting accuracy. Specifically, the models exhibit enhanced predictions for both building energy demand and indoor air temperature profiles over a 12-h forecast horizon at 15-min intervals.

The findings highlight the significant potential of transfer learning applications, particularly in scenarios involving new buildings or recently retrofitted buildings where data is limited, yet an accurate model of building energy consumption is essential. The outcomes of this case study led to the following conclusions.

Despite the differences in electric energy profiles of the whole HVAC system and of the whole building between the source and target buildings, TL still enhances model performance, as evidenced by the positive RRPI values across all thirty-two cases investigated. The performance of the three neural network models (LSTM, MLP, and GRU) improves as more data becomes available for model re-

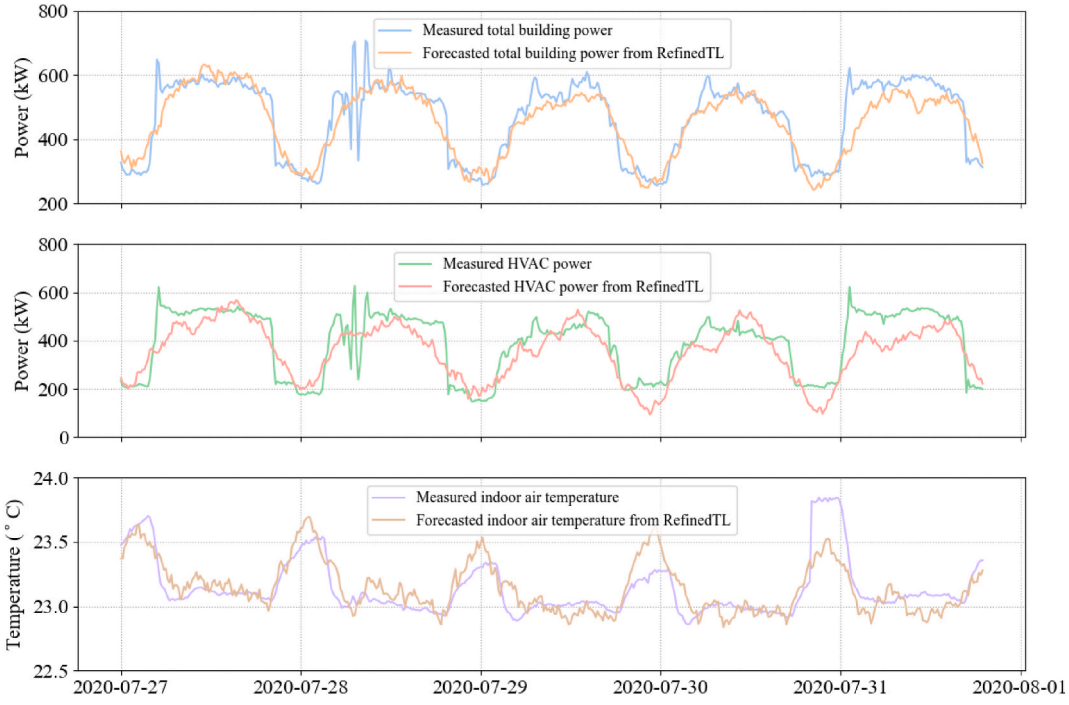

Fig. 11. Comparison of measured and forecasted total building power, HVAC power and indoor air temperature, resulting from the MLP model using the RefinedTL strategy for an example week in the testing dataset of Period 4 of the target building.

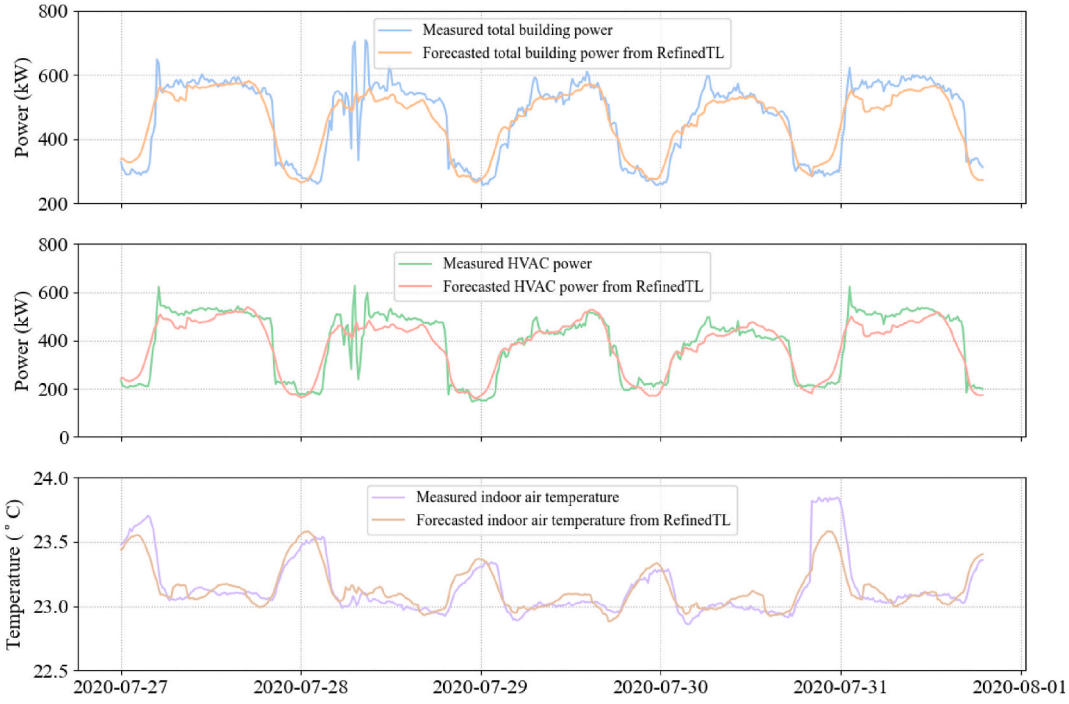

Fig. 12. Comparison of measured and forecasted total building power, HVAC power and indoor air temperature, resulting from the GRU model using RefinedTL strategy for an example week in the testing dataset of Period 4 of the target building.

training, with the best results observed in the MLP model, achieving a CVRMSE of 19 % for P_{HVAC} and 14 % for P_{BLDG}. However, the sensitivity of TL effectiveness varies depending on the selected neural network model, with trends in effectiveness increasing, relatively stabilizing, or decreasing as more data is used for the model re-training on target building. As the training dataset of target building accumulates, the performance of the SelfL strategy gradually improves, eventually surpassing that of DirectTL. In some cases, the addition of just one extra day of data leads to significant improvements in model performance.

This study provides an alternative way to tackle the challenge of data scarcity, which can arise due to limited building operation time, availability of quality data, user privacy concerns, etc., by demonstrating that machine learning models, which typically require large datasets to achieve acceptable accuracy, can benefit from synthetic data. Synthetic data generated from open-source models is often easier to obtain and can significantly enhance model accuracy.

The authors acknowledge several limitations of this work. For instance, the evaluation of thermal comfort could be addressed in greater depth. Future modelling work could focus on a more detailed analysis of thermal comfort, incorporating advanced metrics to evaluate occupant satisfaction. Future work should explore a greater variety of target buildings across different archetypes and climate zones to further evaluate the effectiveness of transferability. Additionally, the control part of the supervisory MPC controller needs to be implemented and tested to evaluate the usefulness of the TL models, and their ability to achieve specific project goals such as enabling precise control of building systems, or activating building energy flexibility. Finally, future work should explore the impact of forecast errors on control performance when the MPC framework is implemented. Since forecast accuracy directly influences the quality of MPC decisions, suboptimal control actions (e.g., overcooling, inefficient equipment scheduling) may be triggered by inaccurate forecasting. These issues could lead to increased energy consumption or reduced occupant comfort.

CRediT authorship contribution statement

Hongwen Dou: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kun Zhang:** Writing – review & editing, Supervision, Software, Resources, Project administration, Funding acquisition, Formal analysis.

Data statement

Due to project confidentiality, the data supporting the findings of this study cannot be shared.

Declaration of competing interest

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

Acknowledgements

The authors acknowledge the support of Natural Sciences and Engineering Research Council of Canada (NSERC) through Dr. Kun Zhang's Discovery Grant RGPIN-2024-03739.

Appendix

The learning curves represented by the variation of mean square error with the number of epochs on both the training dataset and validation dataset of Period 4 using the LSTM model are presented in Figure A1. The optimal number of epochs, determined through a trial-and-error approach, is identified as 50. The discrepancy in mean squared error between the training and validation datasets can be attributed to several factors. First, the model is tasked with forecasting three target variables with distinct scales and variances. Second, there are differences in the data distributions: the training dataset includes data from early July, characterized by relatively lower outside temperatures and building loads, while the validation dataset spans mid-July and end-July when outside temperatures and building loads are higher. These factors contribute to the observed divergence in mean squared error between the two datasets.

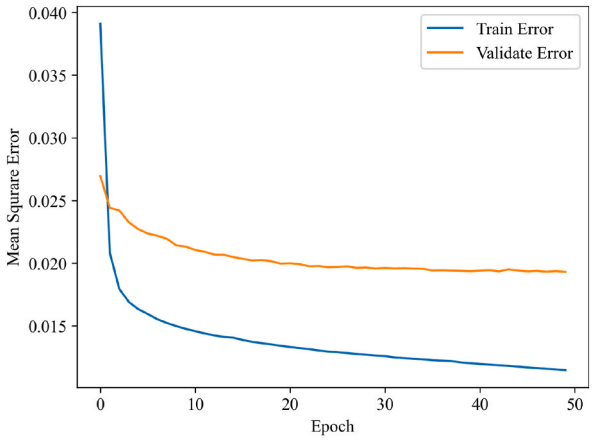

Fig. A1. Learning curve for the training dataset and the validation dataset of Period 4 for the LSTM model.

Figure A2 displays the learning curves for the MLP model, illustrating mean squared error variation with the number of epochs for both the training and validation datasets of Period 4. It indicates the corresponding MLP model is well trained with both training and validation errors decreasing over time and maintaining close proximity.

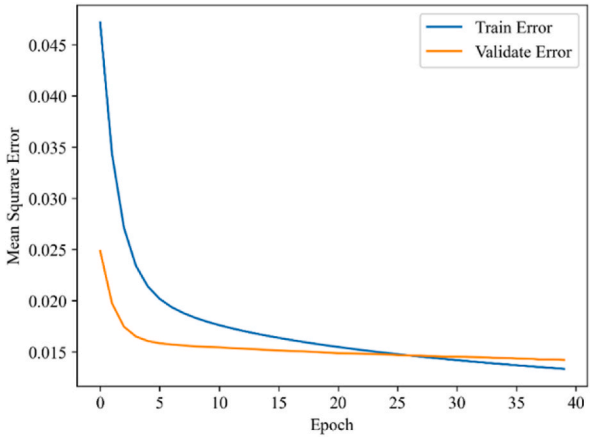

Fig. A2. Learning curve for the training dataset and the validation dataset of Period 4 for the MLP model.

Figure A3 displays the learning curves for the GRU model, illustrating mean squared error variation with the number of epochs for both the training and validation datasets of Period 4. The optimal epoch count, identified through the trial-and-error method, is 300. As training progresses, some differences in mean squared error between the training and validation datasets are observed. However, as discussed above, considering the model's task of forecasting three target variables with varying scales and variances, and the fact that the datasets cover different periods with distinct temperature and building load conditions, these differences should not be interpreted as signs of model overfitting.

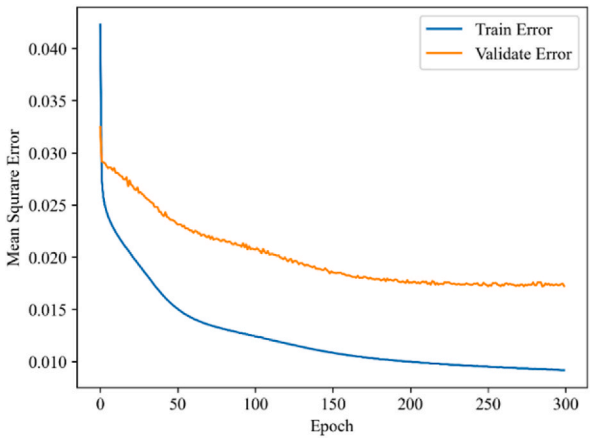

Fig. A3. Learning curve for the training dataset and the validation dataset of Period 4 for the GRU model.

Table A1
Pseudocode outline for transfer learning-based forecasting workflow

Step 1: Preprocess source and target building data
- Handle missing values (e.g., using linear interpolation)
- Apply time alignment and resampling
- Normalize all input features
Step 2: Pre-train models (MLP, LSTM, GRU) on source building data
- Train each model using the training dataset of the source building
Step 3: Train baseline models on target building data (SelfL strategy)
- Train each model from scratch using the training dataset of the target building
Step 4: Fine-tune pre-trained source models on target building data (RefinedTL strategy)
- Load each model from Step 2
- Retrain using the training dataset of the target building
Step 5: Evaluate Direct Transfer Learning (DirectTL)
- Use models from Step 2
- Test on the testing dataset of the target building
- Record performance metrics (RMSE, CVRMSE)
Step 6: Evaluate Self Learning (SelfL)
- Use models from Step 3
- Test on the testing dataset of the target building
- Record performance metrics (RMSE, CVRMSE)
Step 7: Evaluate Refined Transfer Learning (RefinedTL)
- Use models from Step 4
- Test on the testing dataset of the target building
- Record performance metrics (RMSE, CVRMSE)
Step 8: Compute the Ratio of Relative Performance Improvement (RRPI) using Equation (8).

Data availability

The data that has been used is confidential.

References

[1] National Resources Canada, Building energy use surveys. https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/showTable.cfm?type=HB&sector=aaa&juris=ca&rn=2&page=0. (Accessed 6 October 2021).
[2] A National Roadmap for Grid-Interactive Efficient Buildings, U.S. Department of Energy - Building Technologies Office, n.d. https://gebroadmap.lbl.gov/ (accessed November 3, 2024).
[3] A. Yahaya, A.B. Owolabi, D. Suh, Enhancing building energy through regularized Bayesian neural networks for precise occupancy detection, J. Build. Eng. 107 (2025) 112777, https://doi.org/10.1016/j.jobe.2025.112777.
[4] G. Gholamibozanjani, M. Farid, Peak load shifting using a price-based control in PCM-enhanced buildings, Sol. Energy 211 (2020) 661–673, https://doi.org/10.1016/j.solener.2020.09.016.
[5] K. Zhang, M. Kummert, Evaluating the impact of thermostat control strategies on the energy flexibility of residential buildings for space heating, Build. Simulat. 14 (2021) 1439–1452, https://doi.org/10.1007/s12273-020-0751-x.
[6] R.Z. Homod, K.S. Gaeid, S.M. Dawood, A. Hatami, K.S. Sahari, Evaluation of energy-saving potential for optimal time response of HVAC control system in smart buildings, Appl. Energy 271 (2020) 115255, https://doi.org/10.1016/j.apenergy.2020.115255.
[7] S. Baldi, C.D. Korkas, M. Lv, E.B. Kosmatopoulos, Automating occupant-building interaction via smart zoning of thermostatic loads: a switched self-tuning approach, Appl. Energy 231 (2018) 1246–1258, https://doi.org/10.1016/j.apenergy.2018.09.188.
[8] C.D. Korkas, S. Baldi, I. Michailidis, E.B. Kosmatopoulos, Occupancy-based demand response and thermal comfort optimization in microgrids with renewable energy sources and energy storage, Appl. Energy 163 (2016) 93–104, https://doi.org/10.1016/j.apenergy.2015.10.140.
[9] G. Chaudhary, H. Johra, L. Georges, B. Austbø, Transfer learning in building dynamics prediction, Energy Build. 330 (2025) 115384, https://doi.org/10.1016/j.enbuild.2025.115384.
[10] F. Qian, W. Gao, Y. Yang, D. Yu, Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption, Energy 193 (2020) 116724, https://doi.org/10.1016/j.energy.2019.116724.
[11] E. Lashgari, D. Liang, U. Maoz, Data augmentation for deep-learning-based electroencephalography, J. Neurosci. Methods 346 (2020) 108885, https://doi.org/10.1016/j.jneumeth.2020.108885.
[12] J. Feng, Y. Li, K. Zhao, Z. Xu, T. Xia, J. Zhang, D. Jin, DeepMM: deep learning based map matching with data augmentation, IEEE Trans. Mobile Comput. 21 (2022) 2372–2384, https://doi.org/10.1109/TMC.2020.3043500.
[13] EnergyPlus. https://bigladdersoftware.com/epx/docs/8-6/engineering-reference/chillers.html#electric-chiller-model-based-on-fluid-temperature-differences. (Accessed 17 May 2021).
[14] Prototype Building Models, Building energy codes program. https://www.energycodes.gov/prototype-building-models. (Accessed 3 November 2024).
[15] Q. Yang, Y. Zhang, W. Dai, S.J. Pan, Transfer Learning, Cambridge University Press, 2020.
[16] G. Pinto, Z. Wang, A. Roy, T. Hong, A. Capozzoli, Transfer learning for smart buildings: a critical review of algorithms, applications, and future perspectives, Adv. Appl. Energy 5 (2022) 100084, https://doi.org/10.1016/j.adapen.2022.100084.
[17] H. Dou, R. Zmeureanu, Transfer learning prediction performance of chillers for neural network models, Energies 16 (2023) 7149, https://doi.org/10.3390/en16207149.
[18] Y. Tang, K. Yang, S. Zhang, Z. Zhang, Photovoltaic power forecasting: a hybrid deep learning model incorporating transfer learning strategy, Renew. Sustain. Energy Rev. 162 (2022) 112473, https://doi.org/10.1016/j.rser.2022.112473.
[19] J. Liu, L. Hou, R. Zhang, X. Sun, Q. Yu, K. Yang, X. Zhang, Explainable fault diagnosis of oil-gas treatment station based on transfer learning, Energy 262 (2023) 125258, https://doi.org/10.1016/j.energy.2022.125258.

[20] R. Janković Babić, A comparison of methods for image classification of cultural heritage using transfer learning for feature extraction, Neural Comput. Appl. 36 (2024) 11699–11709, https://doi.org/10.1007/s00521-023-08764-x.
[21] H. Lu, J. Wu, Y. Ruan, F. Qian, H. Meng, Y. Gao, T. Xu, A multi-source transfer learning model based on LSTM and domain adaptation for building energy prediction, Int. J. Electr. Power Energy Syst. 149 (2023) 109024, https://doi.org/10.1016/j.ijepes.2023.109024.
[22] X. Fang, G. Gong, G. Li, L. Chun, W. Li, P. Peng, A hybrid deep transfer learning strategy for short term cross-building energy prediction, Energy 215 (2021) 119208, https://doi.org/10.1016/j.energy.2020.119208.
[23] C. Miller, F. Meggers, The Building Data Genome Project: an open, public data set from non-residential building electrical meters, Energy Proc. 122 (2017) 439–444, https://doi.org/10.1016/j.egypro.2017.07.400.
[24] A. Li, F. Xiao, C. Fan, M. Hu, Development of an ANN-based building energy model for information-poor buildings using transfer learning, Build. Simulat. 14 (2021) 89–101, https://doi.org/10.1007/s12273-020-0711-5.
[25] A. González-Vidal, J. Mendoza-Bernal, S. Niu, A.F. Skarmeta, H. Song, A transfer learning framework for predictive energy-related scenarios in smart buildings, IEEE Trans. Ind. Appl. 59 (2023) 26–37, https://doi.org/10.1109/TIA.2022.3179222.
[26] Y. Gao, Y. Ruan, C. Fang, S. Yin, Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data, Energy Build. 223 (2020) 110156, https://doi.org/10.1016/j.enbuild.2020.110156.
[27] M. Ribeiro, K. Grolinger, H.F. ElYamany, W.A. Higashino, M.A.M. Capretz, Transfer learning with seasonal and trend adjustment for cross-building energy forecasting, Energy Build. 165 (2018) 352–363, https://doi.org/10.1016/j.enbuild.2018.01.034.
[28] G. Li, Y. Wu, J. Liu, X. Fang, Z. Wang, Performance evaluation of short-term cross-building energy predictions using deep transfer learning strategies, Energy Build. 275 (2022) 112461, https://doi.org/10.1016/j.enbuild.2022.112461.
[29] Y. Yuan, Z. Chen, Z. Wang, Y. Sun, Y. Chen, Attention mechanism-based transfer learning model for day-ahead energy demand forecasting of shopping mall buildings, Energy 270 (2023) 126878, https://doi.org/10.1016/j.energy.2023.126878.
[30] X. Fang, G. Gong, G. Li, L. Chun, P. Peng, W. Li, A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction, Energy Build. 252 (2021) 111435, https://doi.org/10.1016/j.enbuild.2021.111435.
[31] DesignBuilder, DesignBuilder software ltd. https://www.designbuilder.co.uk, 2024. (Accessed 20 August 2024).
[32] Y. Chen, Z. Tong, Y. Zheng, H. Samuelson, L. Norford, Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings, J. Clean. Prod. 254 (2020) 119866, https://doi.org/10.1016/j.jclepro.2019.119866.
[33] E. Saloux, J.A. Candanedo, C. Vallianos, N. Morovat, K. Zhang, From theory to practice: a critical review of model predictive control field implementations in the built environment, Appl. Energy 393 (2025) 126091, https://doi.org/10.1016/j.apenergy.2025.126091.
[34] P.W. O'Callaghan, S.D. Probert, Sol-air temperature, Appl. Energy 3 (1977) 307–311, https://doi.org/10.1016/0306-2619(77)90017-4.
[35] WUFI. https://www.wufi-wiki.com/mediawiki/index.php/Details:HeatTransfer. (Accessed 18 June 2023).
[36] J. Xiong, L. Chen, Y. Zhang, Building energy saving for indoor cooling and heating: mechanism and comparison on temperature difference, Sustainability 15 (2023) 11241, https://doi.org/10.3390/su151411241.
[37] G. Chitalia, M. Pipattanasomporn, V. Garg, S. Rahman, Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks, Appl. Energy 278 (2020) 115410, https://doi.org/10.1016/j.apenergy.2020.115410.
[38] W.H. Batchelder, H. Colonius, E.N. Dzhafarov, J. Myung, New Handbook of Mathematical Psychology: Volume 1, Foundations and Methodology, Cambridge University Press, 2016.
[39] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. https://www.deeplearningbook.org/. (Accessed 18 June 2025).
[40] S. Afzal, B.M. Ziapour, A. Shokri, H. Shakibi, B. Sobhani, Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms, Energy 282 (2023) 128446, https://doi.org/10.1016/j.energy.2023.128446.
[41] A. Al Bataineh, D. Kaur, S.M.J. Jalali, Multi-layer perceptron training optimization using nature inspired computing, IEEE Access 10 (2022) 36963–36977, https://doi.org/10.1109/ACCESS.2022.3164669.
[42] M. Benedetti, V. Cesarotti, V. Introna, J. Serranti, Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: proposal of a new methodology and case study, Appl. Energy 165 (2016) 60–71, https://doi.org/10.1016/j.apenergy.2015.12.066.
[43] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling. https://doi.org/10.48550/arXiv.1412.3555, 2014.
[44] J. Runge, Short-term Forecasting for the Electrical Demand of Heating, Ventilation, and Air Conditioning Systems, 2021, PhD Thesis, Concordia University.
[45] J. Song, G. Xue, Y. Ma, H. Li, Y. Pan, Z. Hao, An indoor temperature prediction framework based on hierarchical attention gated recurrent unit model for energy efficient buildings, IEEE Access 7 (2019) 157268–157283, https://doi.org/10.1109/ACCESS.2019.2950341.
[46] S. Ben Taieb, A. Sorjamaa, G. Bontempi, Multiple-output modeling for multi-step-ahead time series forecasting, Neurocomputing 73 (2010) 1950–1957, https://doi.org/10.1016/j.neucom.2009.11.030.
[47] M. Le Cam, Short-term Forecasting of the Electric Demand of HVAC Systems, Concordia University, 2016. PhD thesis, https://spectrum.library.concordia.ca/id/eprint/981381/. (Accessed 18 June 2025).
[48] G. Van Rossum, The Python library reference, release 3.9.12, https://www.python.org/. (Accessed 31 August 2022).
[49] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Citro Craig, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Y. Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vanhoucke Vincent, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yu Yuan, Xiaoqiang Zheng, TensorFlow: large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/, 2022.
[50] Standard 90.1—Energy Standard for Sites and Buildings except Low-Rise Residential Buildings, ASHRAE, 2022.
[51] E. Saloux, K. Zhang, Data-driven model-based control strategies to improve the cooling performance of commercial and institutional buildings, Buildings 13 (2023) 474, https://doi.org/10.3390/buildings13020474.
[52] Simulation énergétique des bâtiments - Données météo. https://simeb.ca:8443/index_fr.jsp. (Accessed 3 August 2024).
[53] The pandas development team, Pandas. https://doi.org/10.5281/zenodo.3509134, 2020.
[54] Y. Liu, Y. Mu, K. Chen, Y. Li, J. Guo, Daily activity feature selection in smart homes based on Pearson correlation coefficient, Neural Process. Lett. 51 (2022) 1771–1787.
[55] H. Fraihat, A.A. Almbaideen, A. Al-Odienat, B. Al-Naami, R. De Fazio, P. Visconti, Solar radiation forecasting by Pearson correlation using LSTM neural network and ANFIS method: application in the west-Central Jordan, Future Internet 14 (2022) 79, https://doi.org/10.3390/fi14030079.
[56] ASHRAE Guideline 14-2023 – Measurement of Energy, Demand and Water Savings | ASHRAE, https://store.accuristech.com/standards/ashrae-guideline-14-2023-measurement-of-energy-demand-and-water-savings?product_id=2569793 (accessed June 18, 2025).
[57] J. Shinoda, A. Mylonas, O.B. Kazanci, S. Tanabe, B.W. Olesen, Differences in temperature measurement by commercial room temperature sensors: effects of room cooling system, loads, sensor type and position, Energy Build. 231 (2021) 110630, https://doi.org/10.1016/j.enbuild.2020.110630.