Building thermal dynamics modeling with deep transfer learning using a large residential smart thermostat dataset

Engineering Applications of Artificial Intelligence 130 (2024) 107701

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Building thermal dynamics modeling with deep transfer learning using a large residential smart thermostat dataset

Han Li [a,*], Giuseppe Pinto [b,c], Marco Savino Piscitelli [b], Alfonso Capozzoli [b], Tianzhen Hong [a]

[a] Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, 94720, CA, USA
[b] Politecnico di Torino, Department of Energy, TEBE research group, BAEDA Lab, Corso Duca degli Abruzzi, Torino, 10129, TO, Italy
[c] PassiveLogic, 6405 S 3000 E STE 300, Salt Lake City, 84121, UT, USA

ARTICLE INFO

Keywords:
Building thermal dynamics
Data-driven model
Transfer learning
Thermostat dataset

ABSTRACT

Understanding thermal dynamics and obtaining the computational model of residential buildings enable its scaled application in energy retrofits, control optimization and decarbonization. In this paper, we present a deep learning approach to model building thermal dynamics with smart thermostat data collected from residential buildings, with the goal to investigate model generalizability. In the first stage, we developed and compared different Deep Learning architectures including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) models and CNN-LSTM to predict indoor air temperature in a multi-step time horizon. In the second stage, we implemented a Transfer Learning (TL) process, which aims to improve the prediction performance on a new set of buildings (targets), exploiting the knowledge of related or similar buildings (sources). Different TL strategies and source model identification methods were investigated. The study showed that the CNN-LSTM performed the best among the architectures compared, with an average Mean Absolute Error (MAE) of 0.26 °C for one-hour-ahead (twelve 5-min future steps) predictions. Furthermore, the results showed that freezing the LSTM layer and fine-tuning the other layers of the CNN-LSTM achieved the best performance among four TL strategies, which further improved the performance with respect to a machine learning approach by 10%, and proving the effectiveness and generalizability of the proposed approach. A comparison of three different source model identification methods showed that randomly selecting source models constrained by similar building characteristics can provide good TL performance while retaining simplicity comparing with other quantitative source identification methods.

1. Introduction

According to Eckman et al. (2021), buildings contribute to 40% of total energy consumption and 80% of the peak electric demand. The HVAC systems, responsible for Heating, Ventilation, and Air Conditioning, consume up to 50% of the energy used in buildings (Doe, 2015). The increasing concern about greenhouse gas (GHG) emissions resulting from higher building energy consumption has led to a focus on optimizing the operation of existing buildings. This research interest includes ongoing commissioning, predictive controls, and energy management (Garimella et al., 2022). To achieve this optimization, accurate building thermal dynamics models are crucial. These models predict the indoor air temperature trend based on control actions. Thermal dynamics models can be categorized as white-box, grey-box, and black-box models, each characterized by varying levels of complexity, reliance on data, and associated engineering effort.

White-box models are developed based on the first principles governing energy and mass transfer in buildings. They employ a forward approach, utilizing known information about building characteristics, energy systems, occupant behaviors, and weather conditions. However, calibrating white-box models for existing buildings often requires extensive expertise (Fabrizio and Monetti, 2015). On the other hand, grey-box and black-box models adopt an inverse approach, using measured data to identify a model that describes a building's thermal processes. Grey-box models serve as simplified alternatives to the complex white-box models, using simplified parameters to represent the thermal properties of buildings. For example, thermal resistance (R) and thermal capacity (C) are two types of parameters used in reduced-order models, analogous to electric circuits, to describe heat transfer in buildings (Wang and Chen, 2019). When the values of these parameters are determined through regression with measured data, the models are referred to as grey-box or hybrid models. However, grey-box models often rely on simplified assumptions about external and internal loads, as well as HVAC systems. Furthermore, parameter identification in grey-box models requires well-processed temperature, solar radiation, and heat transfer data, making it challenging to incorporate information from features such as HVAC system runtime and occupant motion.

* Corresponding author.
E-mail address: hanli@lbl.gov (H. Li).

https://doi.org/10.1016/j.engappai.2023.107701
Received 4 July 2023; Received in revised form 16 November 2023; Accepted 10 December 2023
Available online 20 December 2023

Black-box models, which do not require prior knowledge about the building, are solely driven by data. Deep Learning (DL) techniques have gained significant popularity in recent years as they offer a powerful means to approximate the nonlinear dynamics of building thermal systems. In a study by Wang et al. (2020), nine machine learning (ML) algorithms were compared for building thermal load prediction. The results revealed that LSTM achieved a mean absolute error (MAE) of less than 0.4 °C for one-hour ahead predictions. Pinto et al. developed LSTM-based thermal dynamics models to enhance energy management in a cluster of buildings using reinforcement learning control (Pinto et al., 2021). In their respective studies, Mtibaa et al. (2020) and Elmaz et al. (2021) explored different model architectures based on LSTM and CNN-LSTM for predicting indoor air temperature. However, it is worth noting that most of the existing research in this domain has focused on developing models tailored to specific buildings, without adequately addressing the generalizability and transferability of these models to buildings with varying characteristics in different climate regions. Additionally, the utilization of real measurements for training and testing these models at scale has been limited in the literature due to the inherent challenges associated with acquiring a substantial amount of data, which is often scarce within building environments. Therefore, there is a need for further exploration and investigation to assess the effectiveness and adaptability of these models across diverse building types and climates, while also considering the practical constraints related to data availability.

Recent studies have investigated the potential of transfer learning to address challenges related to data availability. Transfer Learning (TL) aims to enhance the performance of prediction tasks on new sets of buildings (targets) by leveraging knowledge from related or similar buildings (sources) (Pan and Yang, 2010). TL has gained significant interest in the energy domain. Peirelinck et al. (2022) examined its role in demand response and building control. They found that TL techniques can help leverage existing domain knowledge and human expertise in addition to sparse observational data, achieving improvements that can exceed 30% in a variety of tasks including supervised machine learning and reinforcement learning. Himeur et al. (2022) reviewed the role of TL in energy systems for sustainable smart cities. They highlighted that TL can provide promising solutions to alleviate data shortage and model generalization problems for a variety of problems including load forecasting (Jung et al., 2020), fault detection and diagnosis (Zhu et al., 2021; Chen et al., 2023), thermal comfort prediction (Somu et al., 2021), non-intrusive load monitoring (Li et al., 2023), renewable energy generation, and smart grid energy trading. Pinto et al. (2022b) specifically discussed its applications in smart buildings including: (i) building load prediction (Fan et al., 2022), (ii) occupancy detection and activity recognition (Mosaico et al., 2019), (iii) building dynamics modeling (Pinto et al., 2022a), and (iv) energy systems control (Coraci et al., 2023). They found that despite TL in the smart building domain has gained increasing research since 2015, the interest on this topic is still at the very early stage. And some of the key common challenges include selecting right source building, quantifying the similarity between buildings, thus avoiding negative transfer.

In the context of building thermal dynamic models, Chen et al. (2020) investigated the use of transfer learning on the hidden layers of a multi-layer perceptron to predict building internal temperature and relative humidity. Similarly, Jiang and Lee (2019) transferred a seq-to-seq LSTM model from one building to simulate temperature evolution in a target building with limited data, fine-tuning the model accordingly. Grubinger et al. (2017) proposed an online transfer learning approach that combined a building dynamic model transferred from another building with a model predictive controller (MPC), overcoming data unavailability challenges and achieving better performance compared to a machine learning (ML) approach that did not leverage knowledge from other buildings. Furthermore, Pinto et al. (2022a) attempted to isolate the contribution of each feature in parameter-based transfer learning, emphasizing the importance of selecting the appropriate source building for this process.

Despite these applications, the selection of the source building is highly influenced by the task at hand, the monitoring infrastructure, and the availability of data. Some approaches for selecting the best source building for energy load prediction tasks have been introduced. Fang et al. (2021) proposed using Maximum Mean Discrepancy (MMD) as a similarity index to weigh the importance of a specific source in an ensemble transfer learning methodology. Similarly, Li et al. (2022) employed MMD to quantify the similarity between the source and target buildings of their domain invariant features. Lu et al. (2021) tested various ML models developed on different sources in the target domain, quantifying the similarity between the two domains based on the performance achieved during testing. They evaluated a similarity measurement index (SMI) and selected the building with the lowest SMI as the source for fine-tuning TL. However, these metrics have primarily been used in energy load prediction applications and have been applied to features extracted from neural networks, limiting their interpretability.

1.1. Research gaps and novelty

Despite the opportunity provided by the use of TL to overcome data scarcity and ease the deployment of machine learning in buildings, the reviewed literature presents the following research gaps:

1. Further studies, that leverage real data, are necessary to characterize the effectiveness of transfer learning for building thermal dynamic models.
2. It is still not clear how to handle building similarity when multiple source buildings are available.
3. There is no application of TL using hybrid model architecture for building thermal dynamics in existing literature.
4. Existing literature did not thoroughly compare the performance of different TL strategies for building thermal dynamic models.

In this paper, we used a CNN-LSTM model architecture that is able to perform a multi-step prediction of indoor air temperature on a large residential dataset that spans over different buildings in three U.S. states, three space types and two heating system configurations. The paper also studies the potentials of applying transfer learning even not in the case of data scarcity, identifying how different transfer learning approaches and source buildings impact the performances of the proposed models, comparing them with a traditional machine learning approach. This study was aimed at closing the literature gap and addressed such gaps as follows:

1. A comparison of different neural network architectures for the development of building thermal dynamic models is carried out.
2. Different transfer learning approaches on dozens of real buildings with abundant monitored data are applied, quantifying improvements with respect to traditional ML-based approaches.
3. Different source building selection processes that leverage metadata and similarity metrics are implemented and compared.

The content of the article is organized as follows. Section 2 provides an overview on the use of deep learning algorithms to model building thermal dynamics, transfer learning approaches and metrics to assess their effectiveness. Section 3 describes the employed methodological framework. Section 4 describes the obtained results while Section 5 provides a critical discussion and the concluding remarks of the work.

2. Background

2.1. Deep learning for building thermal dynamics

It is well known that building thermal dynamics are characterized by non-linearity and time-variance (Aliberti et al., 2019). This is due to the combination of heat transfer processes, solar and internal heat gains and HVAC system operations. Despite the modeling of building thermal dynamics still remains a complex task, the higher availability of measured data in the post-occupancy phase of the building is allowing the data-driven approach to be more and more employed. Data-driven models are able to better reflect the actual building thermal dynamics and provide more accurate predictions of building responses considering that they are more simple, easy to formulate and require less parameterization and computation time than detailed forward models (Wang and Chen, 2019). According to Wang and Chen (2019), the literature categorizes data-driven models for building thermal dynamics into three main categories: thermal resistor-capacitor networks (RC models), discrete-time transfer functions (TF models), and artificial intelligence techniques (AI models). Due to their ease of implementation and physical interpretability, RC models are popular in existing literature (Wang et al., 2019). However, RC models are limited in capturing the non-linear behavior of building thermal dynamics. The parameter identification process can be sensitive to model assumptions and data inputs (Cibin et al., 2023). With the advancement and increasing maturity of deep learning algorithms and computing resources, AI models have gained popularity in modeling building thermal dynamics due to their ability to capture nonlinear relationships. These models have been successfully applied to predict heating or cooling supplies in buildings and the evolution of indoor air temperature over time. Among the existing studies, a few used multilayer perceptron (MLP) models (Jin et al., 2019) or time-delay neural networks (Li et al., 2021), while the majority used RNN-based models. Mtibaa et al. proposed an LSTM-based model that could predict the indoor temperature of multiple zones at the same time, which reduced the prediction error by 50% than MLP models (Mtibaa et al., 2020). Xu et al. used LSTM-based model to predict indoor air temperature in a public building (Xu et al., 2019). They introduced an error correction mechanism after the LSTM layers to improve prediction accuracy. Elmaz et al. developed a CNN-LSTM model to predict indoor air temperature, which showed better performance than MLP and standard LSTM models (Elmaz et al., 2021). In general, the prediction of indoor air temperature evolution in a building through an AI model is a task involving different aspects: the need of extracting features from multiple inputs, the need of a sequential modeling, and a multi-step output in the prediction horizon. A brief background on the three aspects are:

- **Feature extraction**: The process of feature extraction involves transforming the initial set of variables into a new set of processed variables that facilitate the learning process of the model. Feature extraction can be divided into two categories: feature reduction and feature enrichment. While CNN is commonly used for feature extraction in computer vision and image processing, it is also widely employed with multivariate time series data using 1D CNN (Lu et al., 2022).
- **Sequential modeling**: Sequential modeling is particularly suitable for making predictions on sequential data, such as audio signals, text streams, or time series data. Recurrent Neural Networks (RNNs) are a popular category of deep learning algorithms that support sequential modeling. LSTM, a type of RNN, utilizes gating mechanisms to control non-linearity and maintain short-term and long-term memory (Hochreiter and Schmidhuber, 1997). LSTM incorporates three gates – input gate, output gate, and forget gate – to capture and handle long-term dependencies, allowing the model to retain or discard information as necessary (Sherstinsky, 2020).
- **Multi-step prediction**: In the case of multi-step prediction, a data-driven model forecasts the target variable for multiple future steps simultaneously. Multi-step predictions of indoor air temperature are particularly useful for optimizing HVAC operations in buildings, enabling the definition of optimal control policies using model-based approaches like Model Predictive Control or model-free techniques like Reinforcement Learning. Depending on the model architecture, multi-step prediction can be classified into two categories: (i) Iterative methods, where the model generates a single-step prediction and uses it as input for the next step prediction iteratively until the desired horizon is reached. (ii) Direct methods, where the model outputs a complete sequence, often referred to as sequence-to-sequence (seq2seq) methods. Common seq2seq techniques include fully-connected methods using linear layers and attention-based methods, which selectively focus on different parts of the input sequence during decoding at each prediction step. Recent studies have shown that fully-connected methods perform well for short-term building thermal predictions (Pinto et al., 2021).

All these aspects are taken into account in the formulation of the prediction problem of the building indoor air temperature evolution referred to the case study after presented. The next section provides instead a brief overview on transfer learning approaches that can be used to share knowledge between different data sources improving model accuracy and at the same time reducing the amount of training data needed for model development.

2.2. Transfer learning

The transfer learning framework is formally defined using the definitions of domain and task (Pan and Yang, 2010; Pinto et al., 2022a). A domain $D = \{X, P(X)\}$ is made up of a feature space and its marginal probability distribution. Similarly, a task $T = \{Y, f(.)\}$ consists of two components, a target variable and an objective predictive function, usually used to approximate the conditional probability $P = (y \parallel x)$ predict the corresponding target variable of a new instance x. Transfer learning aims to improve the performances of the target task T_t in the target domain D_T, exploiting knowledge from the source task T_S in the source domain D_S, where at least one among domains and task differs. The recent advent of deep neural networks has given rise to a new category of transfer learning, specifically called network-based transfer learning. This technique aims to overcome the challenges of traditional deep learning model development, reducing the amount of data needed to train them and increasing their generalizability. It falls under the category of parameter-based transfer learning, where it is assumed that the source and target tasks can share certain parameters or prior distributions of the model's hyperparameters (e.g., neural networks). In this approach, the knowledge acquired from the source task is transferred to another task through the utilization of shared model weights. Network-based transfer learning can be further classified into the following categories:

- **Weight-initialization**: This approach in network-based transfer learning involves utilizing a pre-trained model on source data to initialize the model's weights when training on target data. Subsequently, the model undergoes further fine-tuning on the target data.
- **Feature extraction**: In this network-based transfer learning approach, certain layers of the model are frozen and used as feature extractors. This approach offers advantages such as reduced data requirements for training the model and the ability to leverage data from different domains, which can be adapted to the input/output dimensions and fine-tuned accordingly.
  To evaluate the effectiveness of a transfer learning framework for a regression task (specifically, the prediction problem addressed in this study) and guide the selection of the best combinations of source and target data, various metrics are described in the following section.

2.3. Performance metrics

Various performance metrics have been proposed in the literature to evaluate the impact of transfer learning on regression tasks. These

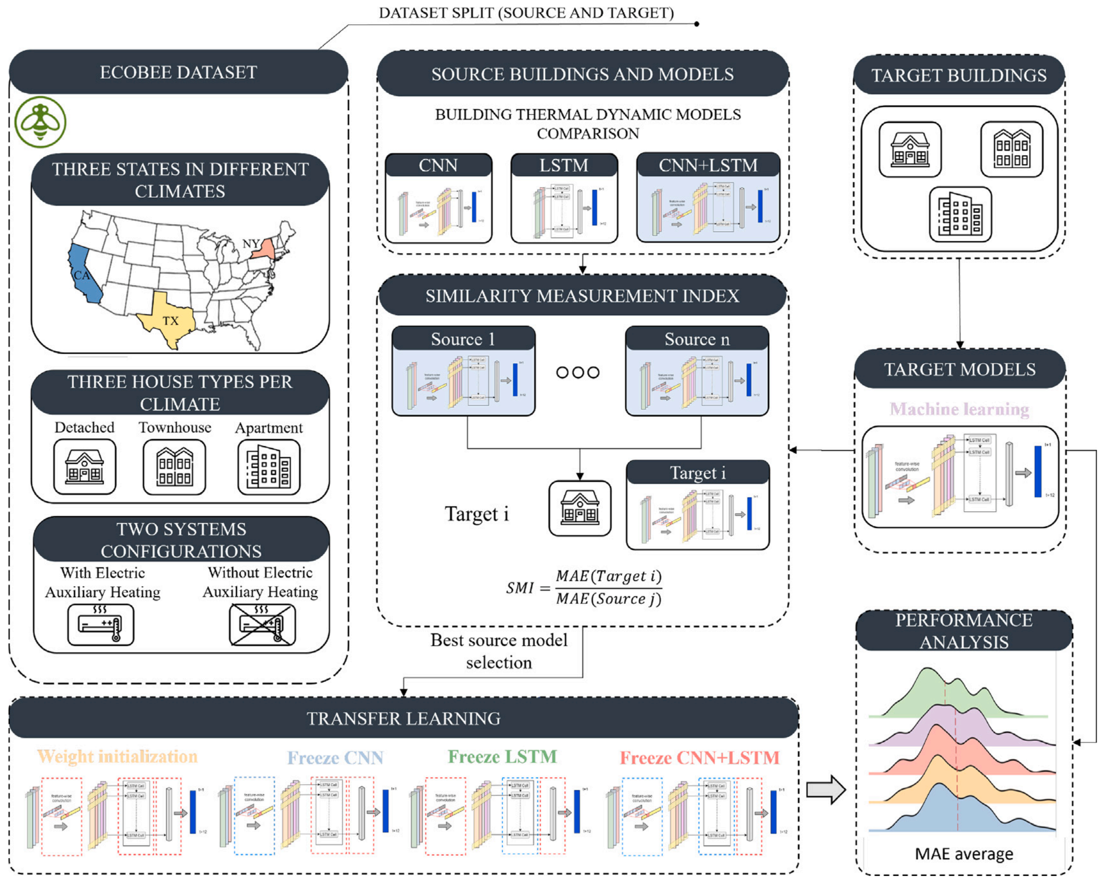


Fig. 1. Overall workflow.

metrics include mean absolute error (MAE), mean absolute percentage error (MAPE), coefficient of determination (R^2), root mean squared error (RMSE), and coefficient of variation of the root mean squared error (CVRMSE). In this study, MAE and R^2 are chosen as the evaluation metrics due to their easy interpretability. Moreover, MAE considers both positive and negative prediction errors, and increase as the error magnitude grows, while R^2 captures the variation in the predictions. Once the performance metric is defined, the performance improvement ratio (PIR) can be computed to measure the relative enhancement in performance (in terms of MAE) achieved by employing a transfer learning framework compared to not using it.

Furthermore, performance metrics such as MAE can be employed to assess whether a regression model developed on a source dataset performs similarly on a target dataset without applying any transfer learning techniques. If the model exhibits similar performance on both datasets, it can be considered as the baseline for evaluating the effectiveness of a transfer learning framework in enhancing performance on the target dataset. In this study, a Similarity Measurement Index (SMI) is utilized for this purpose. The formulas for the metrics employed in the analysis, including MAE, SMI, and PIR, are presented below.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{2}$$

$$SMI = \frac{MAE(Target_{model})}{MAE(Source_{model})} \tag{3}$$

$$PIR = \frac{MAE_{baseline} - MAE_{new}}{MAE_{baseline}} \times 100\% \tag{4}$$

3. Methodology

3.1. Overview

The proposed methodological framework unfolds over four steps which are represented in Fig. 1. Firstly, we sampled and processed the ecobee smart thermostat data from three U.S. states with the aim to retain the distributions of the residential building stock in terms of building and HVAC system types. The sampled dataset was splitted into source and target subsets. Secondly, we developed and compared different deep learning model architectures and evaluated their performances using the source dataset. As a third step, we implemented four different TL strategies based on the best model architecture evaluated within step 2 using the target dataset. Eventually, we compared and analyzed the TL and DL model performance. Details about the four steps are presented in this and following sections.

3.2. Data processing

The smart thermostat data obtained from ecobee's Donate Your Data (DYD) program was utilized. As of 2022, this dataset consisted of over 190,000 households in the United States and Canada that voluntarily shared their data anonymously for research purposes. Each thermostat provided user-reported metadata about the building, including details such as location (at the city level), space type, gross floor area, number of floors, and the time of initial thermostat connection. Several previous studies have explored the dataset and tried to explore the building operation patterns, such as heating and cooling habits (Meier et al., 2019), thermal preferences (Huchuk et al., 2018), and occupancy schedule (Jung et al., 2023). Fig. 2 visually depicts the three steps involved in the data processing. To evaluate the deep learning model's

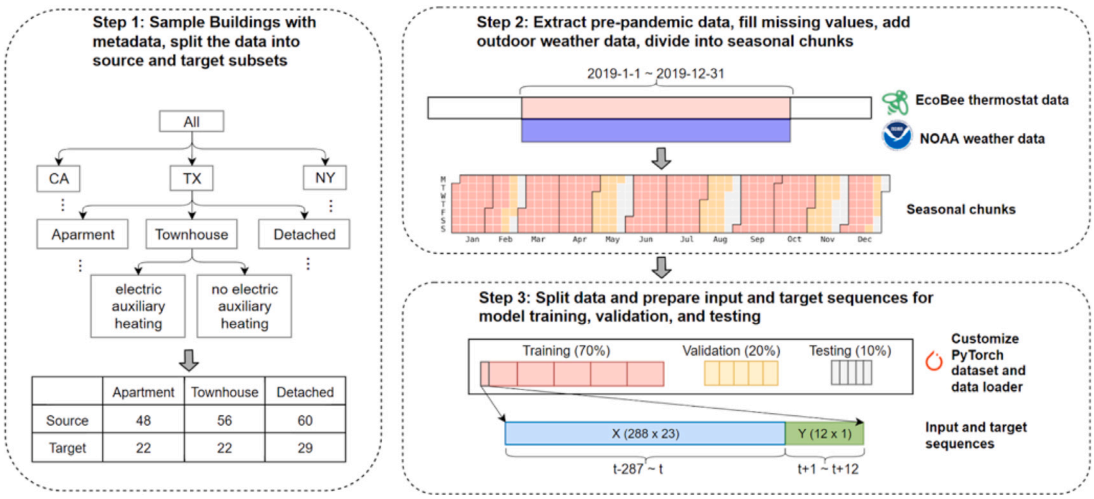


Fig. 2. Data processing.

applicability across buildings with diverse characteristics and in different climate regions, a subset of buildings was randomly selected based on the building metadata. This subset encompassed buildings of three space types (apartment, townhouse, and detached single-family houses) and two HVAC system configurations (with and without electric auxiliary heating) from three distinct climate regions in the United States (California, Texas, and New York). The selected buildings were then divided into a source subset and a target subset, which were used for developing the deep learning (DL) model and the transfer learning (TL) model, respectively.

During the second step, the time-series data for each building was processed. Considering the impact of the COVID-19 pandemic, particularly on occupancy patterns in residential buildings, only data from the entire year of 2019 (from January 1, 2019, to December 31, 2019) was used for testing the proposed methodology. The original time-series data included indoor air temperature and humidity, cooling and heating indoor air temperature setpoint, supply fan runtime, cooling and heating system runtime, and occupant motion detections, with a temporal resolution of five minutes. Temporal features, such as the time of day and the day of the week, have a significant correlation with residential building thermal loads, which typically vary between weekdays and weekends. To incorporate these features into the time-series data, they were encoded as cosine and sine values with corresponding periods. Additionally, holidays were encoded as a binary feature. Outdoor weather conditions are another crucial factor influencing building thermal loads. Since the ecobee thermostat dataset did not include outdoor weather data, outdoor air temperature data was added to each thermostat's information. The latitude and longitude of each thermostat were used to determine the closest weather station listed by the National Oceanic and Atmospheric Administration (NOAA). Sub-hourly outdoor temperature data from that weather station was then retrieved using the NOAA's API and synchronized with the other time-series data. After incorporating the outdoor temperature and temporal features, the dataset consisted of 23 features, which were standardized by scaling to unit variance using the scikit-learn package (Pedregosa et al., 2011). Table 1 provides the names, units, and types of the variables included in the available dataset.

As shown in Fig. 2, the input data has the shape of (B × L × W), where B is the batch size whose value is determined in hyperparameter tuning described in Section 4.1. L is the length of the sequences, which is 288 (24-hour lookback with 5-minute timesteps). W is the width of the sequences or the number of input features shown in Table 1. Two subsequent input sequences are shifted by one timestep (i.e., the feature values at t-287 in the former sequence become the values at t-288 in the current sequence). The dataset is implemented in a custom PyTorch dataset class to handle the sanity check and iterative data sampling for model training.

3.3. Deep learning model development

Based on the literature review discussed in Section 1, it is evident that LSTM models have exhibited strong performance in modeling building thermal dynamics. In this study, we employed a simple LSTM model as the baseline approach, where the data is directly fed into the LSTM cell, followed by a linear layer for multi-horizon indoor temperature predictions. As an alternative architecture, we introduced a 1D CNN module before the LSTM cell to facilitate feature mixing. Specifically, the 1D CNN module operates solely on a feature-wise basis while preserving the time dimension. Fig. 3 illustrates the CNN + LSTM architecture. Both model architectures were implemented using PyTorch (Paszke et al., 2019). We conducted a comparison of model performance using different sets of hyperparameters, optimizers, and learning rate schedulers, which will be presented in the subsequent section, Section 4.

3.4. Transfer learning strategies

After the creation and comparison of different deep learning architectures (see 4.2), the best architecture (CNN-LSTM) was selected as a starting point to implement different transfer learning strategies, which was further tested on a subset of unseen buildings. As introduced in Section 2.2, depending on which model parameters are frozen during the fine-tuning, there are different transfer learning strategies. To investigate their impact on the prediction performance, four transfer learning strategies are compared:

1. Freeze the CNN: the weights of the CNN are frozen and used as feature-extractor, while the weights of the LSTM and the fully connected layer are initialized and fine-tuned on target data.
2. Freeze the LSTM: the weights of the LSTM are frozen, while the weights of the CNN and the fully connected layer are initialized and fine-tuned on target data.
3. Freeze the CNN-LSTM: both the weights of the CNN and the LSTM are frozen and used as feature-extractor, while the fully connected layer is fine-tuned on target data.
4. Weight initialization: the model is used for initialization purpose only and fine-tuned on target data.

3.5. Source model identification

As reported in Section 2.2 the selection of the source dataset and related source prediction model is an essential task to achieve good results in transfer learning (Afridi et al., 2018). The domain shift

Table 1
Time-series data variables in the ecobee DYD thermostat dataset.

Variable name	Description	Type	Unit
TemperatureExpectedCool	thermostat cooling setpoint	numerical	°C
TemperatureExpectedHeat	thermostat heating setpoint	numerical	°C
Humidity	relative humidity	numerical	%
auxHeat1	auxiliary heating system 1 runtime	numerical	s/5 min
auxHeat2	auxiliary heating system 2 runtime	numerical	s/5 min
auxHeat3	auxiliary heating system 3 runtime	numerical	s/5 min
compCool1	cooling compressor 1 runtime	numerical	s/5 min
compCool2	cooling compressor 1 runtime	numerical	s/5 min
compHeat1	heating compressor 1 runtime	numerical	s/5 min
compHeat2	heating compressor 1 runtime	numerical	s/5 min
fan	supply air fan runtime	numerical	s/5 min
Thermostat_Temperature	aggregated thermostat temperature	numerical	s/5 min
Thermostat_Motion	occupant presence	binary	N.A.
T_out	outdoor air temperature from NOAA	numerical	°C
sin_hour	sine of an hour in a 24-h day	numerical	N.A.
cos_hour	cosine of an hour in a 24-h day	numerical	N.A.
sin_day_of_week	sine of a day in a 7-day week	numerical	N.A.
cos_day_of_week	cosine of a day in a 7-day week	numerical	N.A.
sin_month	sine of a day in a month	numerical	N.A.
cos_month	cosine of a day in a month	numerical	N.A.
sin_week_of_year	sine of a week in a 52-week year	numerical	N.A.
cos_week_of_year	cosine of a week in a 52-week year	numerical	N.A.
is_holiday	whether a day is a holiday	binary	N.A.

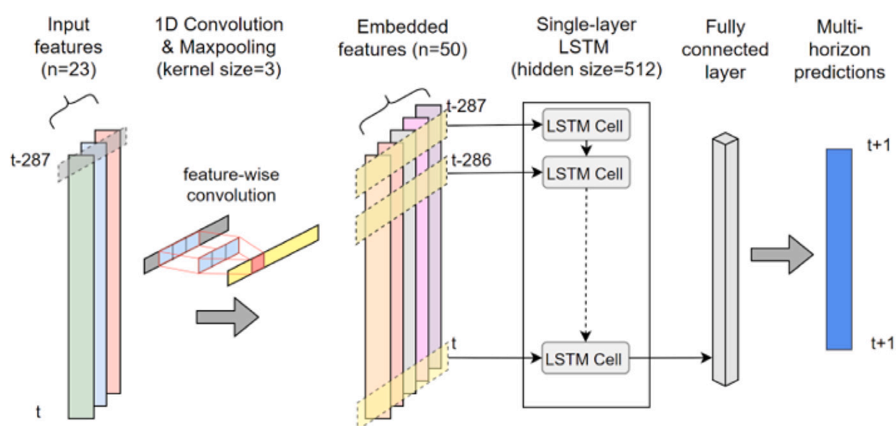


Fig. 3. CNN + LSTM model architecture.

between the source and target domains is a fundamental challenge in transfer learning. The discrepancies between source and target data distributions due to different building characteristics, weather and occupancy patterns can severely decrease the model's performance. A properly selected source model can reduce the amount of data required to effectively fine-tune its parameters on the target dataset and achieve higher values of accuracy for the transferred model respect to a prediction model directly fitted on the target dataset (i.e., positive transfer). On the opposite, an improperly selected source model can even lead to negative transfer, then making the transfer learning process performing worse than a prediction model directly fitted on the target dataset. Depending on the problem, different source identification methods and metrics have been proposed and tested. In general, source models could be selected manually with domain expertise, or automatically with the help of similarity metrics.

- Manual Selection: Manual selection of source models involves leveraging domain expertise to identify models that are particularly well-suited for the target tasks. This approach is valuable when there is deep domain knowledge available, enabling the selection of models that capture relevant features and nuances specific to the tasks under consideration. Manual selection ensures that the source models are aligned with the intricacies of the target domain, potentially leading to superior performance in transfer learning scenarios. For example, in a recently published TL study on building thermal dynamics prediction, the authors selected the source models according to the climate zone and energy efficiency level of the source and target buildings on which they were developed and then fine-tuned through TL strategies (Pinto et al., 2022a).
- Automatic selection: On the other hand, automatic selection methods, guided by similarity metrics, are instrumental in scenarios where the dataset is vast and diverse. Similarity metrics allow us to quantify the similarity of the underlying distributions between different building datasets (Larsson et al., 2021). By employing these metrics, we can systematically assess which source buildings' thermal behaviors are more similar to the target building, thereby facilitating the TL. This data-driven approach helps improve objectivity and efficiency, particularly in situations where manual assessment might be challenging due to the complexity or size of the data.

Despite the existing studies, there is no previous research that investigated how, and to what degree, the source model selection impact the TL performance for DL models trained with large-scale smart thermostat data. Therefore, we investigated different source selection approaches and compared their corresponding target model performance. Specifically, we compared three source selection approaches

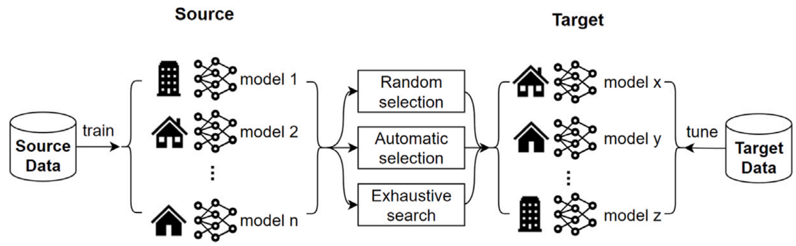


Fig. 4. CNN + LSTM model architecture.

shown Fig. 4: (1) random selection based on similar building characteristics, (2) automatic selection using MMD, and (3) exhaustive search between all possible pairs of source model and target data.

For the random selection, we first manually divide the source buildings into groups based on the available metadata from the smart thermostat dataset, which include locations (i.e., California, Texas, New York), building types (i.e., detached houses, apartments, townhouses), HVAC systems (i.e., whether there is electric auxiliary heating). Then, for each target building, a source model is randomly selected from the corresponding group with the same metadata. For the MMD-guided selection, in addition to the metadata constraint, a source model is identified where the source–target pair has the lowest value of Maximum Mean Discrepancy (MMD) calculated with the thermostat data. Eq. (5) denotes the calculation of MMD.

$$MMD(\mathcal{S},\mathcal{T}) = \left\| \frac{1}{n}\sum_{i=1}^{n}\phi(x_i) - \frac{1}{m}\sum_{j=1}^{m}\phi(y_j) \right\|_{\mathcal{H}}^{2} \tag{5}$$

Here, $\mathcal{S}$ and $\mathcal{T}$ are data distributions with source building and target building samples $x_1,\ldots,x_n$ and $y_1,\ldots,y_m$. ϕ is a feature mapping function that maps the input space to a reproducing kernel Hilbert space $\mathcal{H}$. In this study, a Gaussian kernel is used. The norm in the equation is the norm in $\mathcal{H}$.

Lastly, for the exhaustive search approach, we first test each source model on each target building dataset, always imposing that the source and target buildings should have the same metadata. Then, the source model that achieves the lowest MAE value when directly applied on the thermostat data of a target building is selected as a best starting point for testing TL strategies.

4. Experiment and results

4.1. Model training

As outlined in Section 3, the initial phase of the analysis involves dividing the EcoBee dataset into two distinct groups: source and target buildings. The source dataset consists of a total of 164 buildings, comprising 48 apartments, 56 townhouses, and 60 single-family houses located in California, Texas, and New York. On the other hand, the target dataset encompasses 77 buildings, also situated in the aforementioned three U.S. states. The subsequent step focuses on training thermal dynamics models using each source building, thereby creating a collection of source models for testing various transfer learning (TL) strategies on the group of target buildings.

To ensure optimal training performance and results, the settings and hyperparameters employed play a crucial role. To identify the most suitable hyperparameters, we utilized the Optuna hyperparameter optimization framework (Akiba et al., 2019). The objective was to discover the hyperparameters that minimize the average mean absolute error (MAE) on a randomly selected subset of 10 homes from the source building dataset. The specific hyperparameters available for tuning and their corresponding search space were specified. The optimization process started by randomly sampling from the search space and progressively improving using an evolutionary optimization approach. For both machine learning (ML) and TL model training, we employed the Adam optimizer and a cosine annealing scheduler to gradually reduce the learning rate. The loss function used was the mean squared error (MSE), measuring the disparity between predictions and ground truths. The hyperparameter search space and training configurations for ML are presented in Table 2. Our investigation revealed that the learning rate, CNN kernel size, and LSTM hidden size significantly influenced model performance, while the impact of other hyperparameters was marginal. Consequently, we selected a single-layer LSTM model without dropout for the sequential model.

The models were trained using an NVIDIA Titan RTX graphics card with 24 GB graphics RAM. To optimize training efficiency, we implemented mixed precision training with half precision floating point numbers, enabled by PyTorch's automatic mixed precision (AMP) package. This approach resulted in a notable 30% speedup compared to full-precision training. Each model required approximately 7 min to complete the training process. Since the CNN feature extraction step was relatively straightforward, we did not observe significant differences in training time between the vanilla LSTM and CNN-LSTM models.

For the transfer learning phase, a smaller learning rate was selected to update the weights of the neural network layers compared to the learning rate used during the initial training of the source models. We employed a cosine annealing approach, gradually reducing the learning rate between lr_transfer values of 2e-4 and 2e-5. Furthermore, the source model was selected based on metadata, specifically aiming to identify the most similar building in terms of state, building type, and the similarity measurement index mentioned earlier. By comparing the performance of the transfer learning models with a machine learning model trained solely on target data from scratch, we aimed to assess whether leveraging knowledge from similar buildings could enhance model performance.

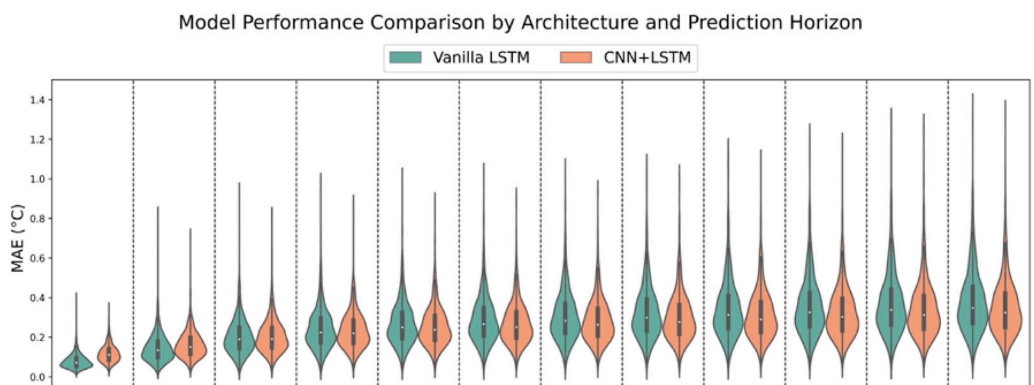
4.2. Machine learning results

The evaluation of the machine learning models developed using the source dataset was conducted using the 10% test data discussed in Section 3.2. The source dataset comprised 48 apartments, 56 townhouses, and 60 single-family houses located in California, Texas, and New York. The performance of the machine learning models was assessed from two perspectives: (1) a comparison of the overall performance between the vanilla LSTM and CNN-LSTM models, and (2) the prediction accuracy of the CNN-LSTM models across different seasons, building locations, types, and HVAC system configurations.

Fig. 5 presents a comparison of the Mean Absolute Error (MAE) distribution for different prediction horizons between the vanilla LSTM and CNN-LSTM models. It is observed that, except for the first three prediction steps (t+1 to t+3), the CNN-LSTM models achieved a lower average MAE compared to the vanilla LSTM models. Overall, there was a 6.6% improvement in performance for all prediction steps with the CNN-LSTM models. Additionally, the standard deviation of MAE for the CNN-LSTM models was generally lower than that of the vanilla LSTM models, except for the first prediction step (t+1). This indicates that the CNN-LSTM models exhibited more consistent performance across most prediction steps. A further comparison of the R^2 values between prediction and ground truth of each future step of the two

Table 2
Hyperparameter search space and selected values.

Hyperparameter	Distribution	Range	Selected
learning rate	log uniform	[2e−4, 2e−2]	2e−3
Adam optimizer weight decay	log uniform	[1e−6, 1e−4]	1e−5
Conv1D kernel size (CNN-LSTM only)	discrete with step=32	[32, 256]	50
LSTM number of layers	discrete with step=1	[1, 4]	1
LSTM hidden size	discrete with step=128	[128, 1024]	512
LSTM dropout probability	discrete with step=0.1	[0, 0.8]	0
batch size	discrete with step=128	[128, 1024]	512
number of epochs	N.A.	N.A.	60



Horizon	t+1	t+2	t+3	t+4	t+5	t+6	t+7	t+8	t+9	t+10	t+11	t+12	All
Vanilla LSTM (mean, std)	(0.088, 0.028)	(0.143, 0.057)	(0.199, 0.07)	(0.237, 0.075)	(0.261, 0.079)	(0.281, 0.083)	(0.299, 0.088)	(0.316, 0.094)	(0.332, 0.098)	(0.346, 0.102)	(0.36, 0.106)	(0.371, 0.109)	(0.271, 0.122)
CNN+LSTM (mean, std)	(0.106, 0.033)	(0.159, 0.052)	(0.201, 0.062)	(0.23, 0.066)	(0.248, 0.069)	(0.263, 0.072)	(0.278, 0.076)	(0.292, 0.081)	(0.306, 0.085)	(0.318, 0.089)	(0.331, 0.092)	(0.343, 0.096)	(0.253, 0.1)
PIR	-20.5%	-11.2%	-1.0%	3.0%	5.0%	6.4%	7.0%	7.6%	7.8%	8.1%	8.1%	7.5%	6.6%

Fig. 5. Vanilla LSTM vs CNN-LSTM MAE comparison by prediction horizon.

architectures has shown a similar trend in Fig. A.1 in Appendix. In summary, the CNN-LSTM models outperformed the vanilla LSTM models, particularly for longer-term predictions. The better performance can be attributed to the feature extraction capability of the CNN layer. Compared with vanilla LSTM, the CNN-LSTM architecture can automatically learn relevant features when the data contains spatial pattern or local dependencies. Moreover, CNNs are known for their robustness to local distortions in the input data. The robustness can ensure that the learned features are more stable across different inputs, making the model more generalizable. Therefore, we selected the CNN-LSTM architecture for the transfer learning task.

In addition, we examine the consistency of CNN-LSTM models. Fig. 6 provides a breakdown of the performance of the CNN-LSTM model (average MAE for the entire prediction horizon) on the source buildings across different seasons, locations, building types, and the presence of electric auxiliary heating in the building. The figure reveals that most of the models achieved an average MAE of less than 0.5 °C, excluding for one apartment in Texas during the spring season, equipped with auxiliary electric heating, which exhibited an average MAE of 0.8 °C. The results indicate that the proposed CNN-LSTM model architecture generalizes well, because there were no significant variations observed between seasons and building characteristics for most models. The performance of the model can be considered satisfactory for applications such as thermal load prediction and optimal control (Elmaz et al., 2021).

4.3. Transfer learning results and comparison with machine learning

The first step to assess transfer learning performance was to evaluate the MAE for each of the 77 target buildings, averaged over the prediction horizon. Fig. 7 shows the distribution for each of the proposed neural networks, with a dashed red line that represents the mean value, ordered from the lowest to the highest mean value. As can be seen, freezing the LSTM layer while fine-tuning the other layers leads to a performance improvement with respect to ML, while the other TL techniques achieve no improvement over standard ML when the amount of data is sufficiently large (e.g., 1 year). The details of the performance metrics of the different TL strategies compared with ML can be found in Table A.1 in Appendix.

After having identified that freezing LSTM layer as the best TL approach, the analysis continued to assess the impacts of different source building selection approaches on the TL performance, following the method described in Section 3.5. Fig. 8 shows the test set MAE distributions of the ML models and TL models with different source selection approaches for 77 target buildings in three U.S. states, with the freezing LSTM layer option. The ML models are trained from scratch with target building data. We conducted t-tests on the three performance metrics to investigate whether the different source model selection methods are significantly better than each other. Not surprisingly, TL with exhaustive source search and MMD both achieved a lower MAE than random selection. However, as shown in Table 3, the improvements are not statistically significant enough to justify the additional computational needs involved in the source building selection process. Therefore, we proceeded with random source building selection. Further t-tests between the TL with random source building selection and ML showed statistically significant performance improvements.

With the TL models fine-tuned with randomly selected source buildings, the analysis continued to quantify the improvement beyond ML for each target model, measured by the MAE improvement. MAE improvement is evaluated as the difference between TL performance and ML ones, therefore, if it is greater than 0, TL performs better, if smaller than 0, ML outperforms TL. Fig. 9 shows the MAE improvement at different aggregation levels: by state, by building type and by system configuration. In each subplot, the scatters above the dashed lines are buildings where TL outperformed ML and vice versa. The higher the scatters, the more improvements. The figure shows the worst performances in New York, followed by Texas and California. Furthermore,

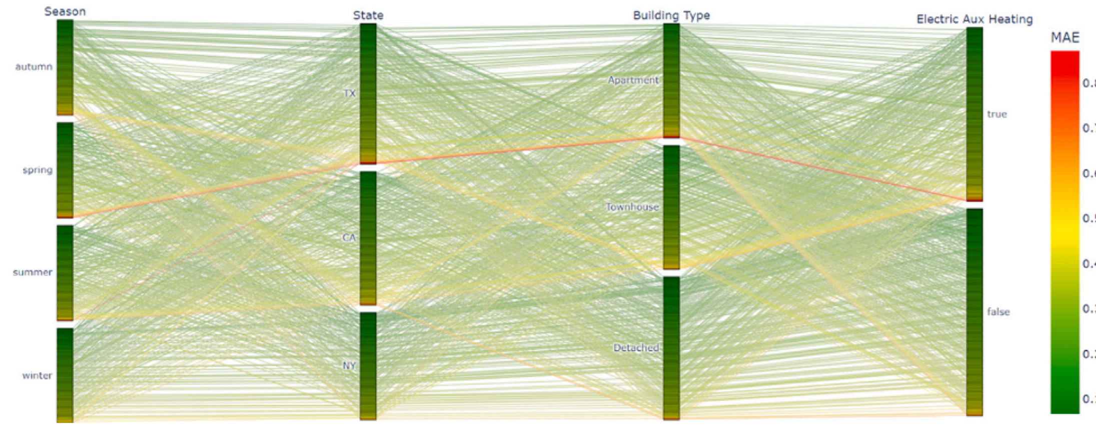


Fig. 6. CNN-LSTM model performance breakdown.

Table 3
P-values for R^2, MAE, and MSE comparisons between different TL strategies.

Group A	Group B	p-score (R^2)	p-score (MAE)	p-score (MSE)
TL - Random	TL - Exhaustive	1.000	0.123	0.662
TL - Random	TL - MMD	0.931	0.820	0.624
TL - Exhaustive	TL - MMD	0.933	0.079	0.359
TL - Random	ML	1.5e−17	1.6e−17	8.3e−17

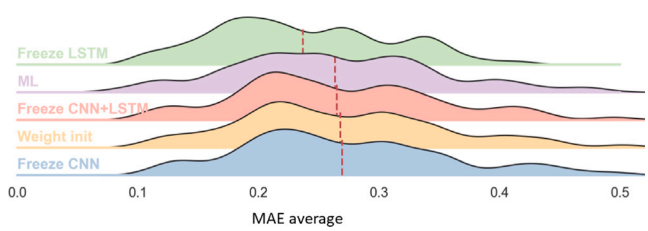


Fig. 7. MAE distribution of different Tl strategies.

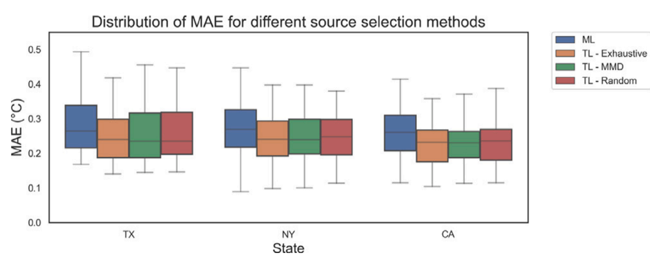


Fig. 8. MAE distribution of different source model selection methods.

looking at building type, the detached houses have lower performances, which could be attributed to their more complex building structures and thermal dynamics, since the higher variability of construction materials might increase the heterogeneity. On the other hand, there is no clear relationship with the system configuration and the TL performance.

Furthermore, to classify the effectiveness of TL and characterize the drivers behind the performance, the following work highlights three particular areas: negative transfer learning, neutral transfer and positive transfer. Neutral transfer occurs when MAE improvement ranges between −0.005 and 0.005 °C, selected as a threshold and representing 2% of the average MAE of machine learning models. Negative transfer occurs when MAE improvement is smaller than −0.005 °C, while positive transfer improves the MAE by at least 0.005 °C.

These classifications were then inserted in a classification tree (see Fig. 10) that used one-hot-encoded variables of state, building type and system configuration, trying to explain in which cases negative and neutral transfer happens. In particular, as previously stated, neutral and negative transfer mainly happens in NY state with detached houses and without auxiliary heating, while other negative transfer happens in Texas detached houses, due to the more complex dynamics (see Fig. 10).

Table 4 shows the aggregated performance of each technique, displaying the mean, median and standard deviation of MAE over the 77 buildings. The models with frozen LSTM achieves the best performances, with not only a lower mean value, but also a smaller standard deviation, with an average improvement of around 10% over classical machine learning approaches.

5. Discussion & conclusion

In this research, we introduced a deep learning methodology to forecast indoor air temperature at multiple time horizons using a comprehensive dataset derived from smart thermostat data collected from residential buildings. The dataset employed in this study encompassed 237 buildings situated in three distinct U.S. states, representing a wide range of building types and HVAC system setups. To facilitate efficient time-series forecasting and analysis, we devised a data processing pipeline specifically tailored for the ecobee dataset, offering promising prospects for future applications. Our proposed approach, employing CNN-LSTM models, exhibited remarkable performance in predicting indoor air temperature. With 1-hour-ahead (12-step-ahead) predictions, the models achieved an average mean absolute error (MAE) of 0.25 °C, surpassing the performance of vanilla LSTM models by 6.6%. Furthermore, we conducted a thorough examination of the model's performance across various seasons, building types, locations, and HVAC system configurations. The results highlighted the versatility and adaptability of our proposed models in accurately predicting indoor air temperature across a wide range of residential buildings with diverse characteristics.

Furthermore, we investigated the effectiveness of transfer learning (TL) in the presence of a large volume of real data, leveraging the CNN-LSTM architecture introduced earlier in four different TL settings. The domain shift between the source and target domains is a fundamental challenge in transfer learning. The discrepancies between source and target data distributions can severely decrease the model's

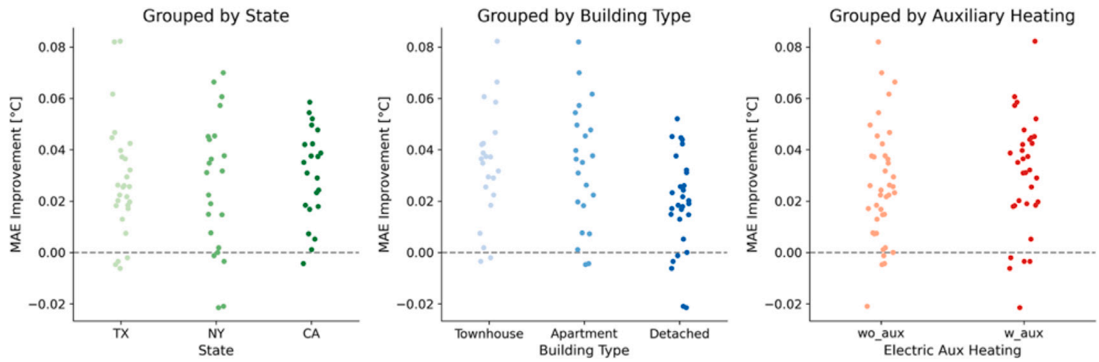


Fig. 9. Performance improvement grouped by different variables: (a) State (b) Building Type (c) System configuration.

Table 4
Aggregated performance of the different techniques.

Strategy	Average MAE [°C]	Median MAE [°C]	Standard deviation [°C^2]
Freeze LSTM	0.237	0.229	0.073
ML	0.263	0.257	0.085
Freeze CNN-LSTM	0.265	0.253	0.083
Weight initialization	0.268	0.259	0.084
Freeze CNN	0.269	0.257	0.085

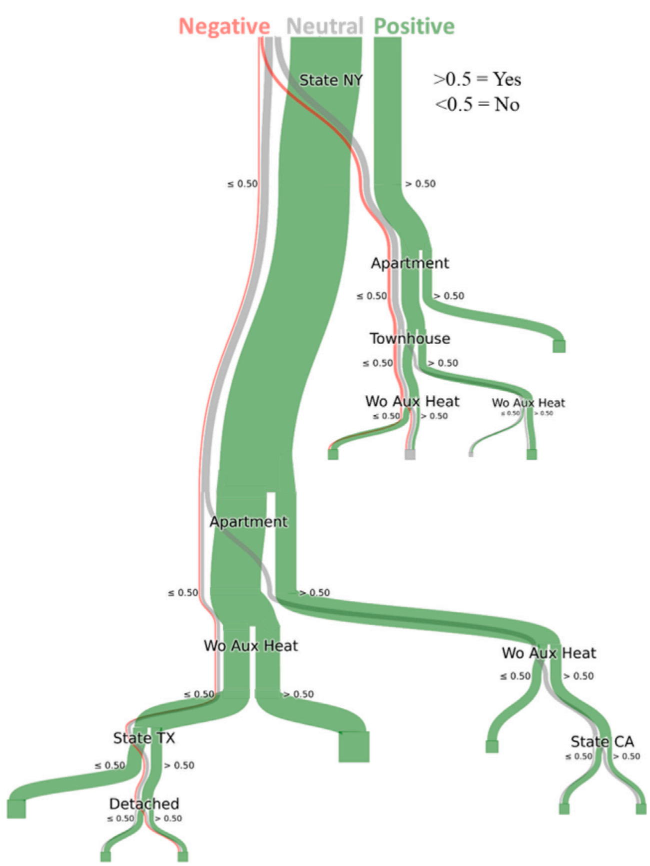


Fig. 10. Classification tree of transfer learning effectiveness according to state, building type and system configuration.

performance. In this study, the potential domain shift was addressed by different measures. Firstly, during the data processing phase, we ensured that the source and target buildings have the same feature space listed in Table 1. When the feature spaces are aligned, the models can learn consistent representations across both domains while keeping the same architecture, making it more likely to generalize well to the target domain. Secondly, our proposed CNN-LSTM model architecture is inherently robust to varying inputs. The convolution layers are capable of extracting key features from the inputs, and the LSTM layers are designed to handle sequential data and capture temporal patterns in diverse domains. Thirdly, during source model training, we applied regularization techniques including dropout and weight decay, which likely prevented the models from being too specialized to the source domain. A crucial aspect of TL effectiveness is the selection of the source building, which can be determined using easily obtainable metadata such as climatic zone, building type, and system configuration. We explored four different techniques that combined feature extraction and weight initialization for the CNN-LSTM model. The results showed that freezing the weights of the LSTM and fine-tuning the CNN and fully connected layers on the target data led to an average improvement of 10%. Moreover, we examined the conditions under which transfer learning performed less effectively, focusing on the influence of climatic conditions and building type on the thermal dynamics of buildings.

We acknowledge that the interpretation of why freezing the LSTM layers achieved the best overall performance requires further investigations. Although the exact reason behind its superior performance needs further investigation, our initial explanations are as follows.

- **Sequential Dependencies**: building thermal dynamics modeling requires capturing intricate temporal patterns and dependencies. LSTMs are specifically designed to model sequences and are excellent at capturing long-term dependencies in sequential data. Therefore, freezing the LSTM layer ensures that the network retains its ability to capture these dependencies, which is crucial for accurate predictions.
- **Feature Extraction**: freezing the LSTM layer preserves the learned sequential patterns, allowing the model to focus on learning the spatial features of the target data set while retaining the learned temporal dynamics from the source task. This focused adaptation likely leads to a more optimal transfer of knowledge.
- **Preventing Catastrophic Forgetting**: freezing the LSTM layer prevents catastrophic forgetting, a phenomenon in which the model forgets important information from the source task when adapting to a new task. Since LSTMs are capable of capturing long-term dependencies, freezing this layer ensures that the knowledge acquired during the source task training is preserved.

The application for TL is often used in data-scarcity contexts. This application demonstrated its effectiveness even in the presence of a

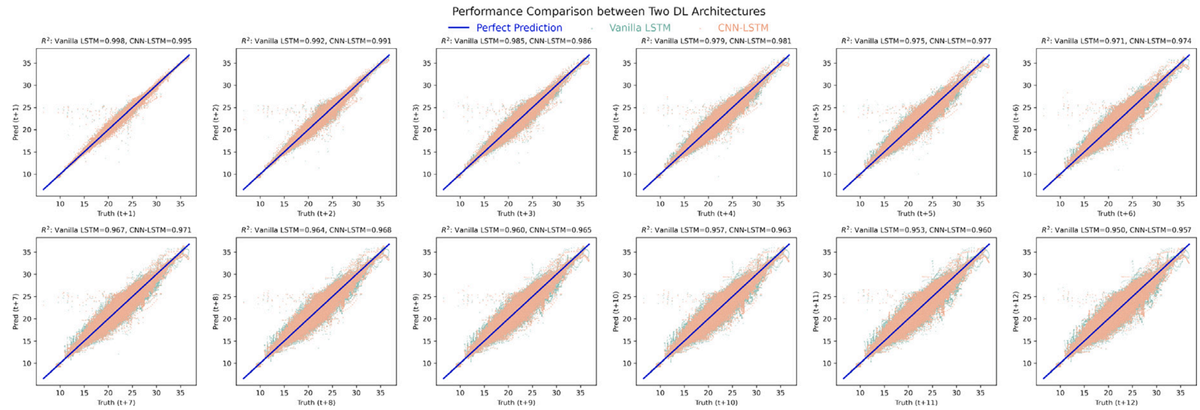


Fig. A.1. Vanilla LSTM vs CNN-LSTM R^2 comparison by prediction horizon.

large amount of residential data, where the stochasticity associated with the occupants is higher relative to commercial buildings. Furthermore, the proposed methodology achieved a performance improvement of 10% with respect to classical machine learning applications. The analysis also suggested that metadata information is effective in selecting the right source buildings, with greater advantages for long prediction horizons. Elaborating on this, future works will investigate the use of an aggregated model, able to incorporate metadata in the inputs to create a single neural network trained on multiple homes used as a source model. The influence of the prediction horizon will be further investigated for a period longer than an hour, exploring embedding and attention schemes in the DNN architecture. Lastly, these models will be coupled in control applications to support the deployment of advanced controllers based on MPC and DRL.

CRediT authorship contribution statement

Han Li: Conceptualization, Study design, Methodology, Writing (lead), Formal analysis, Data curation, Modeling. **Giuseppe Pinto:** Conceptualization, Study design, Methodology, Writing, Formal analysis, Data curation, Modeling. **Marco Savino Piscitelli:** Methodology, Writing – review & edit. **Alfonso Capozzoli:** Conceptualization, Study design, Methodology, Supervision. **Tianzhen Hong:** Conceptualization, Study design, Methodology, Writing – review and edit, Supervision.

Declaration of competing interest

All co-authors declare there is no conflict of interest in the reported work.

Data availability

The authors do not have permission to share data.

Acknowledgments

Li and Hong were supported by the Building Technologies Office of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The work of Marco Savino Piscitelli was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014–2020. We thank ecobee for providing the DYD dataset to support this research.

Appendix. ML & TL performance comparison

Table A.1 shows the performance metrics of the different TL strategies compared with ML.

It can be seen from A.1 that except for the first two future steps, the scatters from the vanilla LSTM models are more sparse than the CNN-LSTM models, corresponding to the lower R^2 values.

Table A.1
Performance metrics for different TL scenarios.

	Mean			Standard deviation		
	MAE (°C)	MSE (°C^2)	R^2	MAE (°C)	MSE (°C^2)	R^2
ML	0.27	0.18	0.79	0.14	0.20	0.42
Freeze CNN	0.28	0.19	0.80	0.13	0.19	0.25
Freeze CNN+LSTM	0.27	0.18	0.80	0.13	0.18	0.25
Freeze LSTM	**0.25**	**0.15**	**0.83**	**0.13**	**0.16**	**0.22**
Weight Initialization	0.28	0.18	0.80	0.13	0.19	0.26

References

Afridi, M.J., Ross, A., Shapiro, E.M., 2018. On automated source selection for transfer learning in convolutional neural networks. Pattern Recognit. 73, 65–75.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2623–2631.

Aliberti, A., Bottaccioli, L., Macii, E., Di Cataldo, S., Acquaviva, A., Patti, E., 2019. A non-linear autoregressive model for indoor air-temperature predictions in smart buildings. Electronics 8 (9), 979.

Chen, Z., O'Neill, Z., Wen, J., Pradhan, O., Yang, T., Lu, X., Lin, G., Miyata, S., Lee, S., Shen, C., Chiosa, R., Piscitelli, M.S., Capozzoli, A., Hengel, F., Kührer, A., Pritoni, M., Liu, W., Clauß, J., Chen, Y., Herr, T., 2023. A review of data-driven fault detection and diagnostics for building HVAC systems. Appl. Energy 339, 121030. http://dx.doi.org/10.1016/j.apenergy.2023.121030, URL https://www.sciencedirect.com/science/article/pii/S030626192300394X.

Chen, Y., Tong, Z., Zheng, Y., Samuelson, H., Norford, L., 2020. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. J. Clean. Prod. 254, 119866.

Cibin, N., Tibo, A., Golmohamadi, H., Skou, A., Albano, M., 2023. Machine learning-based algorithms to estimate thermal dynamics of residential buildings with energy flexibility. J. Build. Eng. 65, 105683.

Coraci, D., Brandi, S., Hong, T., Capozzoli, A., 2023. Online transfer learning strategy for enhancing the scalability and deployment of deep reinforcement learning control in smart buildings. Appl. Energy 333, 120598. http://dx.doi.org/10.1016/j.apenergy.2022.120598, URL https://www.sciencedirect.com/science/article/pii/S0306261922018554.

Doe, U., 2015. An assessment of energy technologies and research opportunities. pp. 12–19, Quadrennial Technology Review. United States Department of Energy.

Eckman, T., Schwartz, L.C., Leventis, G., 2021. Determining utility system value of demand flexibility from grid-interactive efficient buildings.

Elmaz, F., Eyckerman, R., Casteels, W., Latré, S., Hellinckx, P., 2021. CNN-LSTM architecture for predictive indoor temperature modeling. Build. Environ. 206, 108327.

Fabrizio, E., Monetti, V., 2015. Methodologies and advancements in the calibration of building energy models. Energies 8 (4), 2548–2574.

Fan, C., Lei, Y., Sun, Y., Piscitelli, M.S., Chiosa, R., Capozzoli, A., 2022. Data-centric or algorithm-centric: Exploiting the performance of transfer learning for improving building energy predictions in data-scarce context. Energy 240, 122775. http://dx.doi.org/10.1016/j.energy.2021.122775, URL https://www.sciencedirect.com/science/article/pii/S0360544221030243.

Fang, X., Gong, G., Li, G., Chun, L., Peng, P., Li, W., 2021. A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction. Energy Build. 252, 111435. http://dx.doi.org/10.1016/j.enbuild.2021.111435, URL https://www.sciencedirect.com/science/article/pii/S0378778821007192.

Garimella, S., Lockyear, K., Pharis, D., El Chawa, O., Hughes, M.T., Kini, G., 2022. Realistic pathways to decarbonization of building energy systems. Joule 6 (5), 956–971.

Grubinger, T., Chasparis, G.C., Natschläger, T., 2017. Generalized online transfer learning for climate control in residential buildings. Energy Build. 139, 63–71.

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F., Amira, A., 2022. Next-generation energy systems for sustainable smart cities: Roles of transfer learning. Sustainable Cities Soc. 85, 104059. http://dx.doi.org/10.1016/j.scs.2022.104059, URL https://www.sciencedirect.com/science/article/pii/S2210670722003778.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Huchuk, B., O'Brien, W., Sanner, S., 2018. A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data. Build. Environ. 139, 199–210. http://dx.doi.org/10.1016/j.buildenv.2018.05.003, URL https://www.sciencedirect.com/science/article/pii/S0360132318302634.

Jiang, Z., Lee, Y.M., 2019. Deep transfer learning for thermal dynamics modeling in smart buildings. In: 2019 IEEE International Conference on Big Data. Big Data, IEEE, pp. 2033–2037.

Jin, J., Shu, S., Lin, F., 2019. Prediction of indoor air temperature based on deep learning. Sensors Mater. 31.

Jung, S.-M., Park, S., Jung, S.-W., Hwang, E., 2020. Monthly electric load forecasting using transfer learning for smart cities. Sustainability 12 (16), http://dx.doi.org/10.3390/su12166364, URL https://www.mdpi.com/2071-1050/12/16/6364.

Jung, W., Wang, Z., Hong, T., Jazizadeh, F., 2023. Smart thermostat data-driven U.S. residential occupancy schedules and development of a U.S. residential occupancy schedule simulator. Build. Environ. 243, 110628. http://dx.doi.org/10.1016/j.buildenv.2023.110628, URL https://www.sciencedirect.com/science/article/pii/S0360132323006558.

Larsson, H., Taghia, J., Moradi, F., Johnsson, A., 2021. Source selection in transfer learning for improved service performance predictions. In: 2021 IFIP Networking Conference. IFIP Networking, IEEE, pp. 1–9.

Li, X., Han, Z., Zhao, T., Zhang, J., Xue, D., 2021. Modeling for indoor temperature prediction based on time-delay and elman neural network in air conditioning system. J. Build. Eng. 33, 101854.

Li, D., Li, J., Zeng, X., Stankovic, V., Stankovic, L., Xiao, C., Shi, Q., 2023. Transfer learning for multi-objective non-intrusive load monitoring in smart building. Appl. Energy 329, 120223. http://dx.doi.org/10.1016/j.apenergy.2022.120223, URL https://www.sciencedirect.com/science/article/pii/S0306261922014805.

Li, G., Wu, Y., Liu, J., Fang, X., Wang, Z., 2022. Performance evaluation of short-term cross-building energy predictions using deep transfer learning strategies. Energy Build. 275, 112461. http://dx.doi.org/10.1016/j.enbuild.2022.112461, URL https://www.sciencedirect.com/science/article/pii/S0378778822006326.

Lu, C., Li, S., Lu, Z., 2022. Building energy prediction using artificial neural networks: A literature survey. Energy Build. 262, 111718.

Lu, Y., Tian, Z., Zhou, R., Liu, W., 2021. A general transfer learning-based framework for thermal load prediction in regional energy system. Energy 217, 119322. http://dx.doi.org/10.1016/j.energy.2020.119322, URL https://www.sciencedirect.com/science/article/pii/S0360544220324294.

Meier, A., Rainer, L., Daken, A., Ueno, T., Pritoni, M., Baldewicz, D., 2019. What Can Connected Thermostats Tell Us About American Heating and Cooling Habits?

Mosaico, G., Saviozzi, M., Silvestro, F., Bagnasco, A., Vinci, A., 2019. Simplified state space building energy model and transfer learning based occupancy estimation for HVAC optimal control. In: 2019 IEEE 5th International Forum on Research and Technology for Society and Industry. RTSI, pp. 353–358. http://dx.doi.org/10.1109/RTSI.2019.8895544.

Mtibaa, F., Nguyen, K.-K., Azam, M., Papachristou, A., Venne, J.-S., Cheriet, M., 2020. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings. Neural Comput. Appl. 32, 17569–17585.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems, vol. 32.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Peirelinck, T., Kazmi, H., Mbuwir, B.V., Hermans, C., Spiessens, F., Suykens, J., Deconinck, G., 2022. Transfer learning in demand response: A review of algorithms for data-efficient modelling and control. Energy AI 7, 100126. http://dx.doi.org/10.1016/j.egyai.2021.100126, URL https://www.sciencedirect.com/science/article/pii/S2666546821000732.

Pinto, G., Deltetto, D., Capozzoli, A., 2021. Data-driven district energy management with surrogate models and deep reinforcement learning. Appl. Energy 304, 117642.

Pinto, G., Messina, R., Li, H., Hong, T., Piscitelli, M.S., Capozzoli, A., 2022a. Sharing is caring: An extensive analysis of parameter-based transfer learning for the prediction of building thermal dynamics. Energy Build. 276, 112530.

Pinto, G., Wang, Z., Roy, A., Hong, T., Capozzoli, A., 2022b. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. Adv. Appl. Energy 5, 100084. http://dx.doi.org/10.1016/j.adapen.2022.100084, URL https://www.sciencedirect.com/science/article/pii/S2666792422000026.

Sherstinsky, A., 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D 404, 132306.

Somu, N., Sriram, A., Kowli, A., Ramamritham, K., 2021. A hybrid deep transfer learning strategy for thermal comfort prediction in buildings. Build. Environ. 204, 108133. http://dx.doi.org/10.1016/j.buildenv.2021.108133, URL https://www.sciencedirect.com/science/article/pii/S0360132321005345.

Wang, Z., Chen, Y., 2019. Data-driven modeling of building thermal dynamics: Methodology and state of the art. Energy Build. 203, 109405.

Wang, Z., Chen, Y., Li, Y., 2019. Development of RC model for thermal dynamic analysis of buildings through model structure simplification. Energy Build. 195, 51–67.

Wang, Z., Hong, T., Piette, M.A., 2020. Building thermal load prediction through shallow machine learning and deep learning. Appl. Energy 263, 114683.

Xu, C., Chen, H., Wang, J., Guo, Y., Yuan, Y., 2019. Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. Build. Environ. 148, 128–135.

Zhu, X., Chen, K., Anduv, B., Jin, X., Du, Z., 2021. Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. Build. Environ. 200, 107957. http://dx.doi.org/10.1016/j.buildenv.2021.107957, URL https://www.sciencedirect.com/science/article/pii/S0360132321003619.