

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/174049/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zhao, Tianyi, Jiang, Ben, Li, Yu, Rezgui, Yacine , Zhang, Chengyu and Wang, Peng 2025. Multi-point temperature or humidity prediction for office building indoor environment based on CGC-BiLSTM deep neural network. Building and Environment 267 , 112259. 10.1016/j.buildenv.2024.112259

Publishers page: <http://dx.doi.org/10.1016/j.buildenv.2024.112259>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Multi-point temperature or humidity prediction for office building indoor environment based on CGC-BiLSTM deep neural network

Tianyi Zhao^a, Ben Jiang^a, Yu Li^b, Yacine Rezgui^c, Chengyu Zhang^a, Peng Wang^{a*}

a. Institute of Building Energy, Dalian University of Technology, Dalian, China.

b. College of Environment Science and Engineering, Donghua University, Shanghai, China.

c. School of Engineering, Cardiff University, Cardiff, UK.

* Corresponding author: ibewp@dlut.edu.cn

Abstract

The aim of this study is to predict the temperature or humidity changes at multiple relevant points in a building using a deep neural network architecture with multi-task learning to provide more reference information for the design and optimal operation of heating and ventilation systems. For this purpose, traditional multi-task prediction algorithm architecture is combined with Customized Gate Control and other neural networks to build deep neural network architectures for indoor environments with multi-point temperature or humidity prediction tasks. To test the prediction effectiveness of the architecture, a task of predicting temperature or humidity 24h in advance was designed on a real office building indoor environment dataset, and the prediction results were compared with other single-task and multi-task prediction models. Two experimental conditions were designed for this study, one using the complete training set and the other reducing the training set at a certain point. Through the final prediction results, it is found that the multi-task prediction architecture used in this paper shows better or nearly optimal results compared to other prediction models under both working conditions. This study provides some reference value for the application of multi-task prediction algorithms to the task of predicting indoor environments in buildings.

Keywords:

Multi-task prediction algorithm, Deep neural network, Indoor temperature prediction, Indoor humidity prediction, Customized Gate Control

1. Introduction

1.1 Background

At this stage, buildings account for 40% of global energy consumption and a significant 36% of global CO₂ emissions[1]. As a major source of energy consumption in buildings, Heating, Ventilation and Air Conditioning (HVAC) systems have a huge potential for optimization, which will help buildings save energy and reduce carbon emissions. With the rapid development of information technology and modernization, most people spend a lot of time working indoors[2]. Therefore, it is necessary to pay close attention to the changes of building energy consumption and indoor environmental status in order to optimize the operation of the building. A good prediction of the indoor environmental variables can be used as a reference for system regulation to optimize the operation strategy for thermal comfort and energy saving[3]. However, building accurate predictive models remains a challenge due to factors such as data transmission delays and the complex

environment of large buildings.

In indoor environmental regulation studies, the main variables of interest contain temperature and humidity[4, 5]. It is important to control the temperature and humidity of the indoor environment to inhibit mold growth[6]. Inappropriate temperature and humidity levels not only cause many health problems but also have a significant impact on the energy consumption of buildings[7, 8]. Indoor temperature and humidity are usually regulated by controlling the air conditioning system, shading, opening windows and other actions. To optimize control strategies and achieve timely control, it is necessary to understand the trends in temperature and humidity changes in advance, especially for HVAC systems.

1.2 Literature review

Predictions of temperature and humidity are usually made using physical models and data-driven approaches. Temperature and humidity fluctuations are often nonlinear, which is a challenge for the application of physical models. Constructing physical models requires accurate equations to better account for changes in different environmental variables, which is particularly difficult in large public buildings[9]. Another effective way to calculate indoor temperature and humidity is to use simulation software, but it usually requires the collection of many detailed parameters of the building and some expertise in modeling[10]. Data-driven methods are widely used in various time series forecasting tasks due to their simplicity and flexibility[11, 12]. With the advent of the big data era, various sensors installed in buildings effectively collect operational data and also provide certain convenient conditions for model training. Artificial neural networks(ANN) as a method of machine learning are often used to represent nonlinear relationships between variables such as temperature and humidity[13, 14]. It was found that in many cases recurrent neural networks(RNN) perform well in time series prediction such as temperature and humidity[3, 15]. With the increase in computational power and the improvement of network structure on different tasks, deep neural networks based on various recurrent neural networks have been applied in various prediction studies, such as Bi-directional Long Short-Term Memory (BiLSTM)[16, 17], Convolutional Neural Networks and long and short-term memory(CNNLSTM)[18, 19] and the algorithms of Attention and RNN combination[20, 21]. In the study of the prediction of time series such as temperature and humidity, Yang[22] et al, constructed a long and short-term memory model(LSTM) with a feed- forward attentional mechanism for multi-step prediction of temperature and humidity in a solar greenhouse. Its effectiveness was demonstrated by calculating the error ranges at different time steps in advance. Zhen[23] et al. proposed a multi-area indoor temperature prediction method based on LSTM with encoder and decoder architecture and able to predict 2-3 days ahead of time. Furkan[24] et al, by using a CNNLSTM architecture for predicting the indoor temperature of a room on a campus, experimentally proved to be an effective black-box approach for this task by comparing it with Multilayer Perceptron(MLP) and LSTM at different time scales. It can be found that the task of predicting indoor environmental variables can be effectively addressed to varying degrees using data-driven approaches. In addition, some exploratory studies are beginning to be conducted for the prediction of environmental variables at multiple points.

For large office buildings, in order to optimize the control strategies for the air conditioning system and other actions to achieve a more comfortable indoor environment, it is usually necessary to divide the building into multiple zones. For some adjacent areas or open spaces, the indoor temperature and humidity may have different fluctuation ranges but still have a certain degree of influence on each other. Within the same office building, combining the trends of environmental

variables at multiple spatial points with their historical data may impact the accuracy of prediction tasks at different points. Multi-task learning might be one approach to exploring this issue. Multi-task learning is a machine learning approach that aims to learn multiple tasks to improve the generalization ability of a module at the same time[25]. In the process of multi-task learning, the learning of individual tasks is enhanced by sharing information, which in turn improves the accuracy and efficiency of the target task to a certain extent. Previous multitask learning frameworks include hard parameter sharing, cross-stitch[26], sluice-network[27], etc. And multi-task learning has also been studied to some extent on time series prediction tasks, and has been improved and applied to different tasks[28]. Qiao[29] et al, combined the frameworks of deep learning and multi-task learning to explore the deep spatiotemporal features of wind speed data through collaborative learning and multi-site sharing, and obtained good prediction results in advance single-step and multi-step wind speed prediction tasks. Liao[30] et al, constructed a short-term wind probability prediction model based on a multi-task Gaussian process, proving the advantages of the multi-task architecture when there is insufficient data for the target task. Multi-task learning architecture improves learning efficiency through information sharing and can be used to reduce the problem of data sparsity in the target task. In building, multi-point indoor environmental parameter prediction provides more comprehensive reference data for control system optimization and design, and provides the best environmental comfort in the whole management area while reducing energy consumption[9]. At the same time, through the interconnection of environmental variables at multiple points, it may also be possible to achieve a mutually reinforcing effect in order to enhance the overall prediction results. However, when dealing with multiple similar tasks at the same time, there may be positive or negative impacts between the tasks due to features such as data distribution and fluctuation trends in different datasets. How to improve the overall prediction effect and reduce the negative impact on different tasks during training is another important element in dealing with multi-task learning.

Some deep neural networks are sensitive to data distribution, variance, and other factors[31]. The inherent shortcomings of these tasks may compromise the predictive effectiveness of some tasks when model parameters are widely shared[32]. Researchers have begun to make improvements to algorithmic architectures to deal with this problem. These include Multi-gate Mixture of Experts (MMoE) [32] and Customized Gate Control (CGC)[33] networks. These networks are also used in other areas of research[34-37]. Where MMoE handles different tasks through gated networks based on input parameters and the bottom expert network, but ignores the differences and interactions between tasks so it is prone to seesaw phenomena[33]. In the CGC architecture, it contains both task specific unique information and shared common information[38]. Expert tasks and shared tasks are separated at the bottom of the CGC and PLE architectures to reduce the harmful effects between public and task specific knowledge. Chen[39] et al, to further mitigate the seesaw phenomenon, separated the task specific components from the shared components and constructed Temporal Pattern Attention (TPA-LSTM) network, which allows each task to select adaptive information in its favor and achieved better performance on wind and photovoltaic power generation combined prediction task. And there are a number of issues that still need to be explored in multi-point indoor environmental variable prediction tasks.

1) Currently, single-site indoor environmental parameter prediction over different time scales using data-driven methods has become a hot research topic. However, multi-task learning architectures for multi-point prediction tasks with different environmental variables still need to be

further explored and improved.

2) In the task of predicting multi-point indoor environmental variables, it is particularly important to design and utilize multi-task learning architectures appropriately. At the same time, it is necessary to design experiments to deeply explore the performance effects of the architecture.

3) In the multi-point indoor environmental variable prediction task, when the training set of a task is reduced and affects its prediction results, the impact of the multi-task learning architecture on this and other tasks remains to be explored in depth.

1.3 Targets and works

In this paper, we will combine the CGC architecture and BiLSTM to build a multi-task deep neural network to train an indoor temperature or humidity prediction model, and design experiments to explore its performance effect on multiple relevant point prediction tasks. In addition, we also discussed whether this network architecture can improve the prediction performance of the target task while ensuring that other prediction tasks are not greatly negatively affected by reducing the training set of a certain point environment parameter. We will explore the effectiveness of the multi-task prediction algorithm architecture used in this paper by designing comparative experiments with other single and multi-tasks in different prediction tasks. In order to ensure the authenticity and reliability of the whole experiment, real data will be used for model training and validation. The task of forecasting temperature and humidity 24 hours in advance was designed. Hyperparametric optimization of all models to ensure fairness of the experiment. Through the final prediction results, it can be found that the multi-task prediction architecture used in this paper shows better or close to optimal results in all the designed experiments compared to other prediction models.

The following is the architectural design of the chapters of this study. Chapter 2 will focus on describing the main methods used in this study and building the multi-task prediction algorithm architecture for multi-point temperature or humidity prediction. Chapter 3 will provide the main description of data sources, experimental design, and data processing aspects. Chapter 4 will explain the prediction results of the different experiments. Chapter 5 is a discussion of the results. Finally, there are the main conclusions and shortcomings of this study.

2. Methods

2.1 BiLSTM

Recurrent neural networks with gating units have been invented in order to allow the network to selectively memorize or forget some information in a time series task to improve the effectiveness of the overall model training and to reduce the training difficulty to a certain extent. LSTM is one representative RNN.

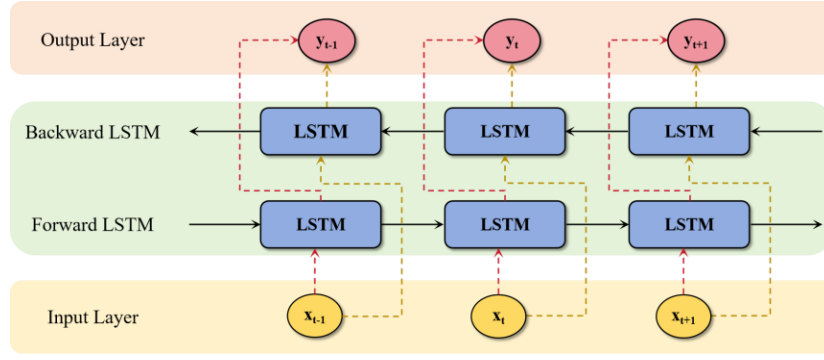


Fig.1. BiLSTM algorithm architecture

In order to take into account both past and future information and thus understand the contextual information in a sequence more comprehensively, BiLSTM was invented. As shown in Fig.1, BiLSTM combines the LSTM in both directions, where the LSTM is applied to the input sequence (namely the forward layer) during the computation process, after which the reverse form of the input sequence is fed into the LSTM (namely the backward layer)[40]. The computation of two LSTMs improves the long-term dependence of the learning and thus improves the accuracy of the model[41]. Therefore, in order to better capture the trends of indoor environmental variables over time, this paper will utilize BiLSTM as the underlying architecture for the deep neural network.

2.2 Multi-task learning

Multi-task learning also generalizes knowledge across different tasks. Learning is performed by utilizing multiple learning tasks that are related to each other to obtain useful information, maximizing the results of each task, and can be used to reduce the problem of data scarcity. According to the objectives and training process of this study, it can be defined as supervised learning with multiple tasks. Each task in the multi-task supervised learning process is based on the prediction of labels from unlabeled data based on labeled training datasets.

Assuming m learning tasks $Y_i (i=1,2,\dots,m)$, where all or some of the tasks are related but not the same, multi-task learning aims to utilize the knowledge in these m tasks to improve the learning performance of all. In this study, we need to simultaneously handle multiple related tasks for temperature or humidity prediction at various points. During the training process, we aim to optimize the prediction performance and reduce negative impacts by capturing the intrinsic connections between different tasks. Since the multi-task processing architecture of CGC includes both unique information for specific tasks and shared common information, and separates these two types of information to mitigate the harmful effects between public knowledge and task-specific knowledge. This research will design a deep neural network architecture for multi-point prediction of indoor environmental variables based on the CGC network architecture.

2.3 CGC-BiLSTM algorithm architecture for multi-points prediction

In order to be able to predict the indoor temperature or humidity at different points using historical data, this paper will build a multi-task prediction algorithm architecture by combining neural networks such as BiLSTM under the CGC architecture. The temperature and humidity datasets from different points of the same building are divided into training and test sets, and the sliding window approach is used to train the model on the training set and test the prediction effect of the model on the test set.

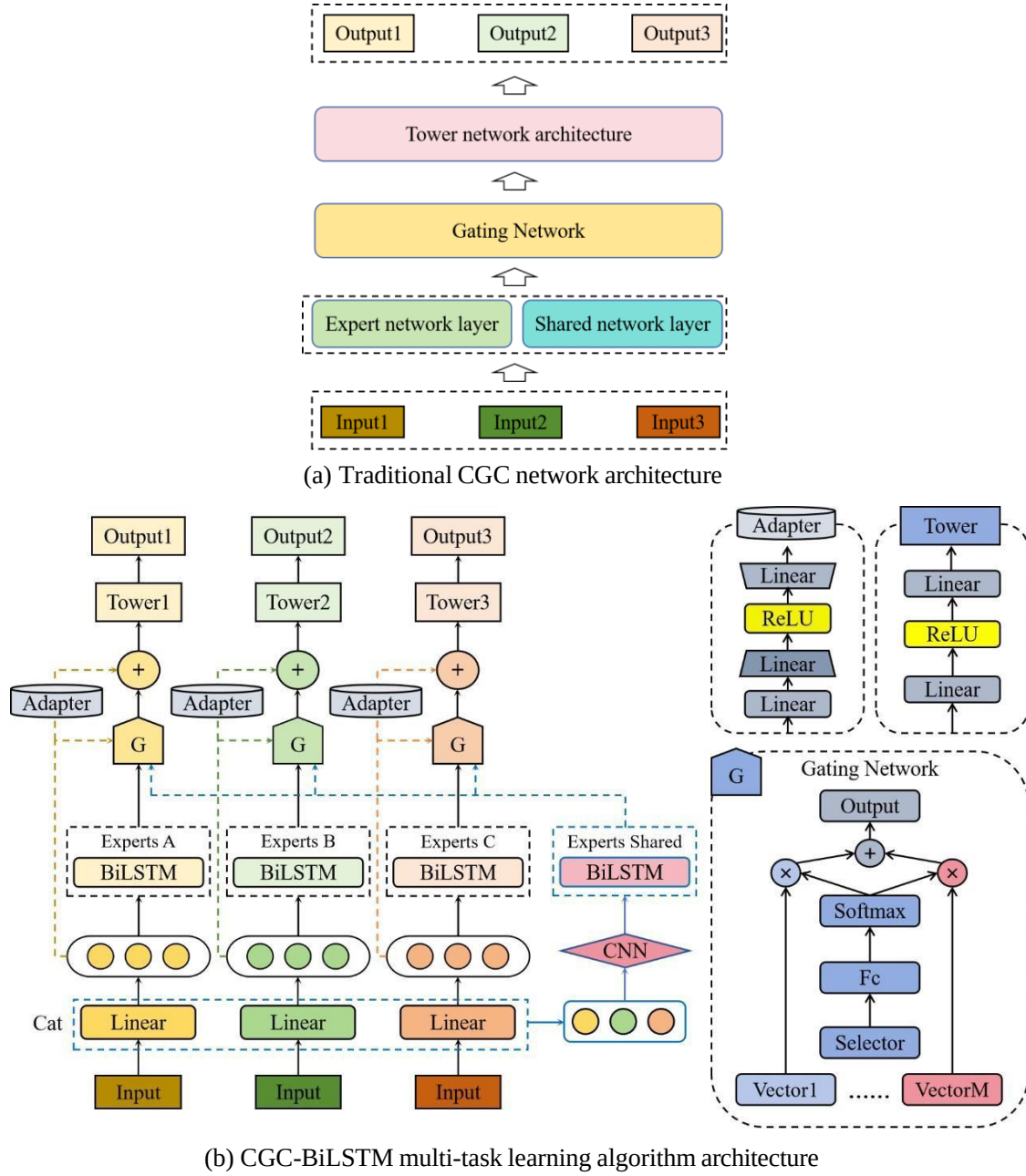


Fig.2. Traditional CGC Network Architecture and CGC BiLSTM Multi Task Learning Network Architecture

As shown in Fig. 2 (a), the traditional CGC architecture consists of individual expert network layers for each task and a shared network layer for all tasks. The gated network computes the information from the shared and expert network layers and feeds it into the tower network layer for computation. Finally, the output of the network is obtained. The multi-task prediction algorithm architecture(CGC-BiLSTM) used in this study is shown in Fig.2 (b). Firstly, the historical data of each point are input into the corresponding linear layer of the expert network respectively, and then all the outputs are spliced in the parameter dimension into the shared expert network. Reduced training difficulty by feature extraction and dimensionality reduction through convolutional neural networks(CNN) at the shared expert network layer. BiLSTM is used as the bottom network to extract the time features of different point data. A separate gating network was then designed for each task modeled following the CGC network architecture. The weights generated by each gating

network and the features extracted from the task-specific and shared expert layers are computed and summed to serve as inputs to the next layer. Each expert and shared layer in the CGC architecture can be set up with a different number of feature extraction networks. In this paper, the number of feature extraction networks is respectively set to 1 and 2 in each specific expert layer and shared expert layer. Finally the top level features for each task are extracted by a tower network, which is designed to consist of linear layers and activation functions in order to reduce the training difficulty. In order to maintain the stability of the network and to improve the prediction for different tasks, an adaptive module (Adapter) with downward and upward projected linear layers and activation functions is added separately for each task. The architecture is similar to that of the adaptive module mentioned in the article[42]. The multi-task learning algorithm architecture used in this paper combines the features of CGC and separates task public and task specific parameters, reducing the harmful interference between public and task specific knowledge. The input parameters of different tasks are also separated at the input layer to further reduce the computational effort of the network and the negative impact between different tasks.

However, when training the model using a multi-task architecture, the losses from different tasks will be obtained, so the final loss function formula needs to be determined. We will draw on the automatically weighted loss function equation[43]. The multiple tasks are defined as $Y_T(Y_{T1}, Y_{T2}, \dots)$. To optimize the learnable parameter w , a loss function for multi-task learning is defined as $L_T(x, y'_T, y_T; w_T)$, where y'_T and y_T denote the predicted and true values, respectively. In turn, each single-task loss function is defined as $L_t(x, y'_t, y_t; w_t)$. The original article calculates the final loss function formula by adding a learning coefficient c to account for differences between single-task losses and a regularization term as shown below.

$$L_T(x, y'_T, y_T; w) = \sum_{t \in T} \frac{1}{2c_t^2} \cdot L_t(x, y'_t, y_t; w_t) + \ln(1 + c^2) \quad (1)$$

For ease of calculation, c is set to the same fixed value in this study. In this study mean square error loss (MSE) is used as the loss function L_t for each single task and is calculated as follows.

$$MSE = \frac{\sum_{i=1}^n (y_t - y'_t)^2}{n} \quad (2)$$

Where n is the number of samples, y_t is the actual value, and y'_t is the predicted value. The closer the MSE is to 0 indicates that the difference between the predicted value and the real value is smaller. Adam will be used as the optimizer of the network during the network training process, and the maximum number of iterations will be set to further prevent overfitting. In order to ensure the fairness and validity of the comparison results in subsequent experiments, this loss function is also used for other multi-task prediction models.

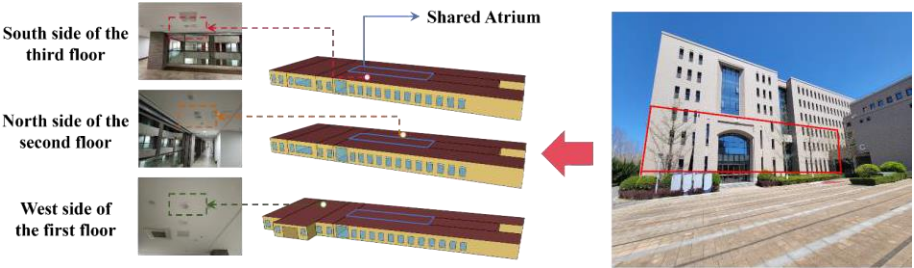
3. Multi-point temperature or humidity prediction for indoor environment case study

3.1 Data source

The goal of this study is to combine the CGC architecture and BiLSTM to build a multi-task deep neural network for indoor temperature or humidity prediction tasks at multiple relevant points. In this study, two experiments were designed including the experiment with the complete training set and the experiment in which the single training set was reduced and affected the prediction results. The experiments are compared with several common single-task prediction models as well as multi-task prediction models to explore the performance effects of the models used in this study. All

experiments will be performed with real datasets.

In this study, indoor temperature and humidity datasets were collected from an office building of a university in Dalian, Liaoning Province, China. The dataset is sourced from the 1st to 3rd floors of the building, with the collection period ranging from March 13, 2023, to December 31, 2023, and a collection interval of 1 hour. The building is primarily used from 9:00 AM to 6:00 PM, and its main purpose in the collected floors is for office work. The building is cooled by split air conditioners in summer and heated by a centralized heating system in winter.



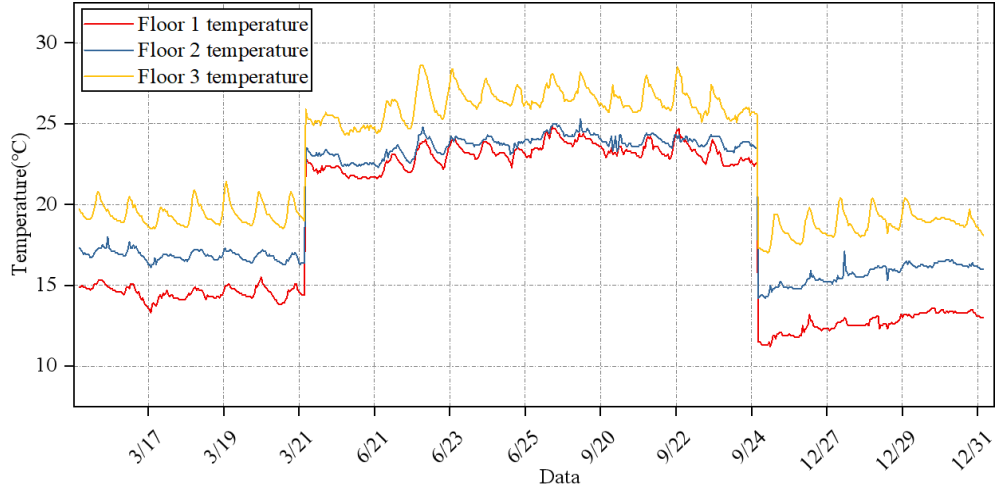
(a) Sensor installation distribution diagram

Sensor appearance	Equipment parameters
	Equipment type: iBest-REM-4W-1 Precision target: Temperature: $\pm 0.2^{\circ}\text{C}$ Humidity: $\pm 2\%\text{RH}$ Installation method: Ceiling or wall mounted Working environment: Temperature: $-30\text{--}60^{\circ}\text{C}$ Humidity: $0\text{--}99\%\text{RH}$

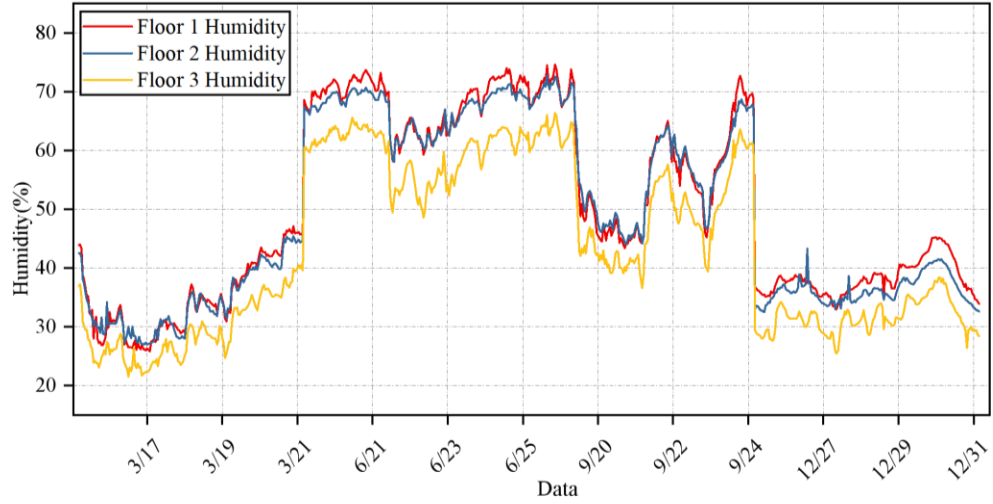
(b) Parameters related to environmental monitoring sensors

Fig.3. Sensor installation location and information

Environmental sensors have been placed in the corridors of floors 1-3 to monitor temperature and humidity changes in real time. As shown in Fig. 3(a), the sensor installation locations are in the west corridor of the first-floor atrium, the north corridor of the second-floor atrium, and the south corridor of the third-floor atrium. The relevant parameters of the sensors are shown in Fig. 3(b). As illustrated in Fig. 4(a) and (b) represent the real-time monitoring values of temperature and humidity by sensors on different floors (Floor1, Floor2, Floor3) during different time periods. It can be observed that due to factors such as the sensor installation locations, different floors, and other indoor activities, the temperature and humidity fluctuation ranges at different monitoring points exhibit certain differences.



(a) Temperature fluctuations on different floors



(b) Humidity fluctuations on different floors

Fig.4. Real-time indoor temperature and humidity monitoring results

The atrium of the building connects floors 1-3, and there may be some correlation between environmental variables at different points due to heat exchange and other factors. Therefore, this study will explore the performance effect of multi-task model in indoor environmental variable prediction task based on this dataset. It can be observed that due to the different locations of the observation points, although the overall trends of temperature and humidity fluctuations at different observation points during the same time period are similar, there are some differences in the fluctuation ranges. It is evident that predicting the temperature or humidity at different points within the same building that have certain correlations is a similar task. However, influenced by spatial and other related factors, there are differences in the distribution of temperature and humidity data at different observation points.

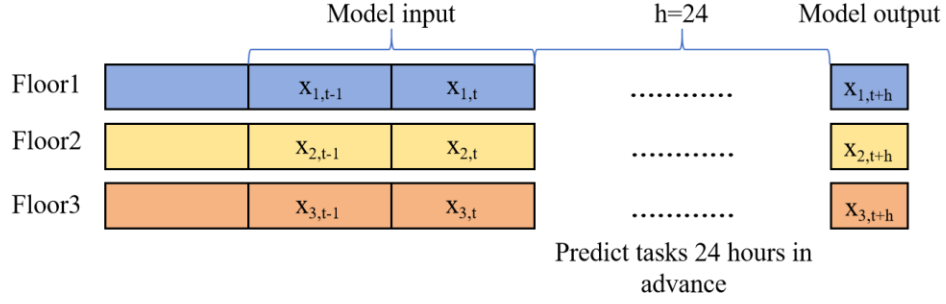


Fig.5. Application of sliding windows

As shown in Fig. 5, the building temperature and humidity prediction experiment for 24 hours in advance was designed in this study. The sliding window method was used to select the historical data, and the window size was set to 2. Specifically, the target task of this study is to predict the trend of indoor temperature or humidity after 24 hours using historical data consisting of the current time and the previous time of the hour. The dataset used in this study is relatively complete, with only a small amount of missing data, which may be due to errors in the collection process or sensor malfunctions, and which we will fill in using nonzero data averages from before and after time points. For the convenience of training, while considering that only a hourly environmental parameter dataset of less than a year has been collected. Therefore, the dataset was sorted by collection time and the first 7000 sets of data were used to construct the experimental dataset. The first 60% is used to build the training set(The model output date is March 14 to September 5), and the remaining 40% is used to build the test set(The model output date is September 6 to December 31). This classification method may overlook the impact of factors such as seasonality and external environmental changes on model training, which also increases the difficulty of model training to a certain extent.

3.2. Experimental design

The main objective of this study is to construct a model architecture suitable for temperature or humidity prediction at several related points using a multi-task learning architecture and deep neural networks. At the same time, explore whether the architecture can be used to improve the prediction performance of a target point when its training set is small, and reduce the impact on the prediction results of other points. To achieve this goal, we will conduct a series of comparative tests to verify the validity of the model as shown in Fig.6.

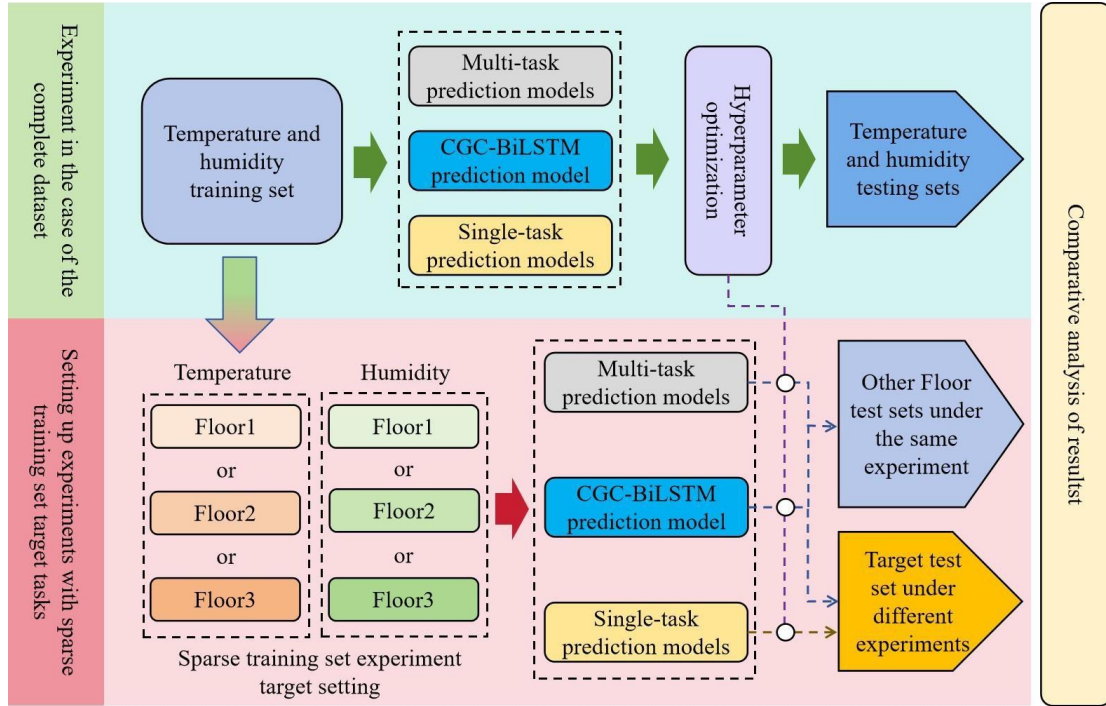


Fig.6. Experimental design and procedure

(1) In terms of the full dataset, we will train the multi-task prediction model used in this paper as well as the comparison model on the full dataset of different points (floors). First, we train the single-task prediction model for each point location separately and compare it with the performance of the multi-task prediction model (CGC-BiLSTM) used in this study on the test set. We will explore whether the predictive power of the CGC-BiLSTM prediction model is close to or better than that of the single-task prediction model (BiLSTM, All-BiLSTM, LSTM, CNNLSTM) at each point. Subsequently, we will introduce other common multi-task prediction models (Share Bottom, MMoE, and CGC) as comparative models to explore whether the algorithmic architecture used in this study optimizes the overall prediction results to some extent.

(2) Due to reasons such as sensor installation time, there may be situations where insufficient data is collected from a particular point, which may affect the prediction of the model. We test the prediction ability of CGC-BiLSTM for the target task and other tasks by reducing the temperature or humidity training set at different points and setting it as the target task. This is done by replacing a certain amount of continuous data at the earliest collection time of the training set with 0 to simulate the situation of insufficient training data in the earlier stage. Because of the need to keep the amount of data consistent across all tasks during multi-task prediction model training, we filled the rest of the target task training set with zeros. During testing, we observed the effect of model performance on all tasks when setting each of the three points as a target task and reducing the number of training sets. Depending on the difficulty of the task, we will set the target task training set for temperature and humidity with different amounts of data. Through initial observations, we found that humidity fluctuates more dramatically than temperature, which may make humidity more difficult to predict accurately and will require a relatively high number of datasets. The target task training sets for temperature and humidity were set to only the last 1500 groups (close to half the number of test sets) and 3000 groups (close to the number of test sets), by experimental testing, respectively.

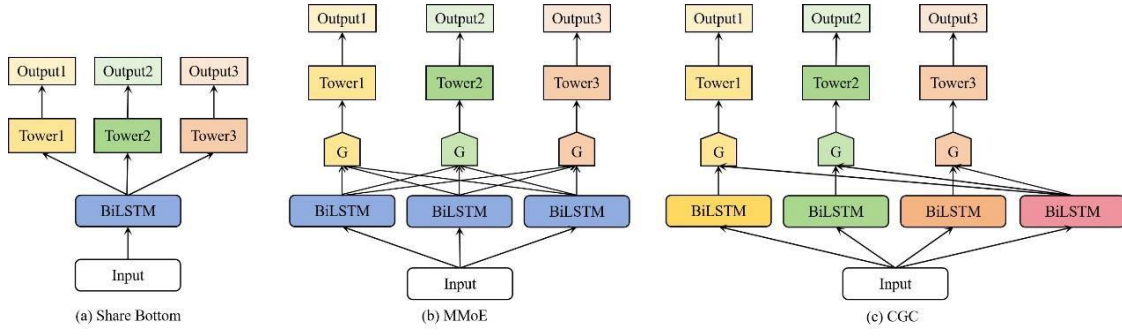


Fig.7. Multi-Task Model Training Architecture for Comparison

Some of the single-task prediction architectures used include BiLSTM, CNLSTM, LSTM, and a model that inputs all point history parameters into BiLSTM at the same time (All-BiLSTM). The multi-task predictive model used for the comparison will be trained using some traditional architectures (Share Bottom, MMoE, and CGC), and these architectures are shown in Fig.7. To ensure the fairness of the experiments, we replaced all the underlying neural networks used for feature extraction with the same BiLSTM neural networks as in this study.

3.3 Data processing

Before model training, data need to be normalized or standardized in order to increase the speed and accuracy of model training. In this study, MinMax normalization operation will be performed on the input parameters with the following formula.

$$x_{std} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (3)$$

$$x_{std2} = x_{std} (x_{max} - x_{min}) + x_{min} \quad (4)$$

Where x_{std} represents the normalized data, x_{max} and x_{min} represent the maximum and minimum values of this parameter respectively, and finally the inverse normalization is performed to get the final result x_{std2} .

3.4. Hyperparameter optimization

Hyperparameter optimization is very important for the black-box model, and the determination of each hyperparameter is again a very tedious task for the training of the model[44]. Bayesian optimization will use the performance effects of previously searched parameters to infer the next parameter search as a way to improve search efficiency[45]. Tree-structured Parzen Estimator (TPE) Bayesian optimization is one of the Bayesian optimization methods[46]. This optimization will be used in this study to optimize the hyperparameters that need to be determined for the temperature and humidity prediction models.

The parameters that need to be determined before optimization are shown in Table 1, and the optimization range, optimization objective, and search step size are initially determined by human tuning results. Random seeds are set during the search process to maximize the reproducibility of the optimal results. For CGC-BiLSTM and other multitask prediction models, the BiLSTM layer and its output layer (Linear1), the first linear layer in the tower network (Linear2) as well as the search step size will be optimized. The optimization objectives and search ranges in single-task and multi-task prediction models are shown in the Table 1.

Table 1

Hyperparametric optimization objective and optimization range

Model	Objective	Range	Step
CGC-BiLSTM	BiLSTM size	[300,400]	20
	Linear1	[300,400]	20
	Linear2	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
MMoE	BiLSTM size	[300,400]	20
	Linear1	[300,400]	20
	Linear2	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
CGC	BiLSTM size	[300,400]	20
	Linear1	[300,400]	20
	Linear2	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
Share Bottom	BiLSTM size	[300,400]	20
	Linear1	[300,400]	20
	Linear2	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
BiLSTM	BiLSTM size	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
All-BiLSTM	BiLSTM size	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
CNNLSTM	LSTM size	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001
LSTM	LSTM size	[300,400]	20
	Learning rate	[0.0003,0.0008]	0.0001

The test set R^2 is used as the optimization objective on the single-task prediction model, and the mean value of R^2 on the test set of all tasks is used as the optimization objective on the multi-task prediction model. The number of searches was set to 20, and finally, the optimal hyperparameter combinations of all models for different tasks were obtained.

The formula for R^2 is shown below. $f(x)$ represents the data predicted by the model, y represents the real data, and n is the number of data. The closer R^2 to 1 indicates that the model fits the observed data better.

$$R^2 = 1 - \frac{\left(\sum_{i=1}^n (f(x_i) - y_i)^2 \right)}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right)} \quad (5)$$

3.5 Evaluation indicators

In order to evaluate the prediction results more effectively, this paper uses the coefficient of variation measuring the root mean square error (CV-RMSE) as an evaluation index, in addition to choosing the more commonly used mean absolute error (MAE) and R^2 as the evaluation indexes in this paper[47]. The above evaluation indicators are calculated as follows, where n is the total number of observations, p_i is the predicted value, a_i is the i th true value, and $\bar{a}$ is the average value.

$$CV - RMSE = \frac{1}{n-1} \sqrt{\sum_{i=1}^n (a_i - p_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - p_i| \quad (7)$$

4. Discussion of results

4.1. Temperature prediction results under the complete data set

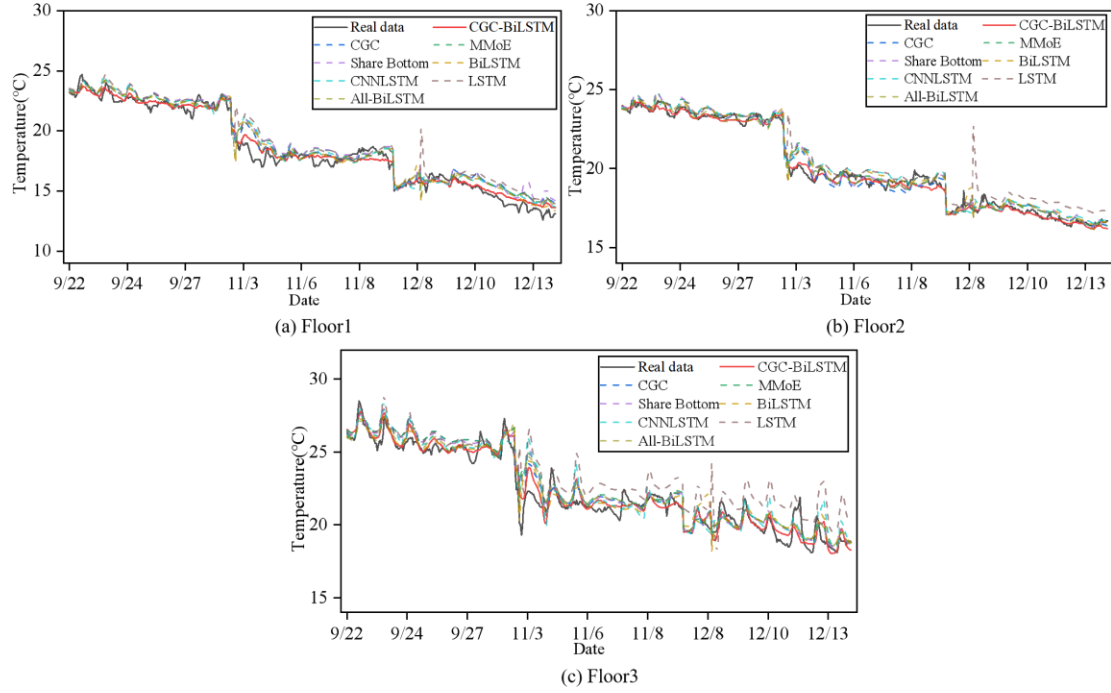


Fig.8. Indoor temperature prediction results under complete training set

Due to the large test set, in order to better observe the trend of the prediction results at different points in time, we divided the whole test set into 3 equal parts and screened the prediction results for 7 days near the midpoint of each part to show (9.22-9.28, 11.3-11.9, 12.7-12.13). As shown in the Fig.8, although the indoor temperature trends at different points in the same building during the same time period are similar, the range of fluctuations may be somewhat different, especially on the first and third floors. As a whole, it is observed that the prediction results of the CGC-BiLSTM architecture used in this paper are closer to the real data at most of the time points. From the observation of the prediction results on different time periods, the prediction errors of all models are relatively small at the beginning of the test set. In the second time period all the prediction models including the CGC-BiLSTM used in this paper produce large errors near some of the daily extreme points, but the model we use performs better near these points. In the third time period, the model used in this paper can roughly describe the trend of temperature fluctuations, but there is a certain deviation from the real data at some points, which is more significant in other comparative models, especially in the prediction tasks on Floor 3. This situation may be because the model has reached the limit of its predictive capabilities and needs to be trained on the latest historical data to update its parameters. At the same time, since this study divides the indoor temperature and humidity monitoring data for the entire year into training and test sets based on the collection time, the model training process overlooks factors such as seasonal variations in the data and changes in external

environmental conditions. However, this also provide certain challenges to the predictive performance of the model.

Table 2 shows the results of the calculation of the evaluation indicators for different prediction models on the task of temperature prediction at different points, as well as the results of the calculation of the overall mean values. It can be found that all models show good predictive ability due to the small and periodic degree of temperature fluctuation. CGC-BiLSTM performed the best on two of the prediction tasks, and the results were closer to those of two of the single-task prediction models on the other task with a maximum R^2 difference of 0.02. Combining the prediction results of all the tasks, the CGC-BiLSTM multi-task prediction model we used performed best on the multi-point temperature prediction task with an average R^2 of 0.972. The improvement is 3.51% and 0.94% compared to the optimal single-task and multi-task prediction models, respectively.

Table 2

Calculation of evaluation metrics for different prediction models on temperature prediction task

Area	Evaluati on metrics	CGC- BiLST M	CGC	MMoE	Share Botto m	BiLST M	CNNL STM	LSTM	All- BiLST M
Floor1	MAE	0.523	0.783	0.877	0.876	0.540	0.576	0.709	0.537
	CV- RMSE	0.040	0.063	0.075	0.071	0.040	0.042	0.053	0.039
	R^2	0.969	0.924	0.891	0.903	0.969	0.966	0.945	0.971
Floor2	MAE	0.331	0.466	0.610	0.439	0.381	0.409	0.636	0.368
	CV- RMSE	0.022	0.033	0.049	0.029	0.026	0.027	0.044	0.024
	R^2	0.981	0.957	0.909	0.969	0.974	0.971	0.927	0.978
Floor3	MAE	0.454	0.678	0.724	0.610	0.581	0.569	1.213	0.620
	CV- RMSE	0.028	0.040	0.043	0.036	0.036	0.034	0.066	0.036
	R^2	0.967	0.929	0.918	0.944	0.943	0.948	0.806	0.942
Mean	MAE	0.436	0.642	0.737	0.641	0.501	0.518	0.852	0.509
	CV- RMSE	0.030	0.045	0.056	0.045	0.034	0.035	0.054	0.033
	R^2	0.972	0.937	0.906	0.939	0.962	0.962	0.893	0.963

4.2 Humidity prediction results under the complete data set

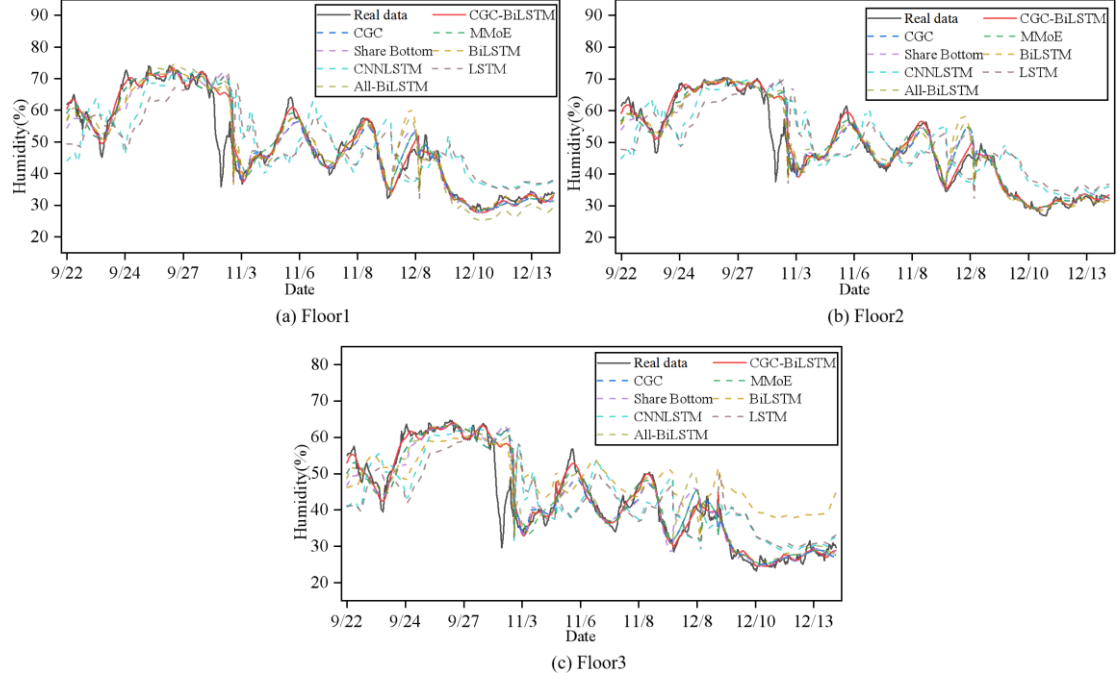


Fig.9. Indoor humidity prediction results under complete training set

Different from temperature, humidity is more affected by the environment and changes more rapidly, which makes it more difficult to be predicted. As shown in Fig.9, the humidity prediction results of different models at different points are shown. It can be found that CGC-BiLSTM performs better in both overall and local prediction effects. Some of the prediction models have significantly deviated from the real data in the last time period, such as BiLSTM and CNNLSTM in the single-task prediction model and Share Bottom in the multi-task prediction model. On the last day of the first time period, we find a drastic change in indoor humidity, which may be due to a sudden change in the weather outside, as the indoor temperature is similarly affected on this day. All the models are too late to cope with this situation, so none of them have a better prediction for this piece of data. In the remaining time periods shown in Fig.9, the multi-task prediction architecture employed in this paper yields more stable humidity prediction results across various locations, with few instances of significant deviation from the actual data.

Table 3 shows the results of the evaluation metrics calculated for different models on the task of predicting indoor humidity at different points. It can be found that the CGC-BiLSTM used in this paper performs comprehensively better on all tasks and the overall results are better than the other compared models. The distributional characteristics of humidity cause some single-task prediction models to perform poorly in the overall results. On the other hand, since the multi-task prediction model captures the characteristics of humidity changes at the relevant points during the training process, the evaluation metrics of the multi-task prediction model are calculated better compared to the single-task model. In the multi-task prediction model, CGC-BiLSTM reduces the negative impact between different prediction tasks by separating the input parameters of different tasks during training and designing an independent expert network for each task. This design architecture affects the final prediction results to some extent. The multi-task prediction model used in this paper improves the average R^2 on different prediction tasks by 4.99% and 7.86% compared to the optimal single-task and multi-task prediction models, respectively.

Table 3

Calculation of evaluation metrics for different prediction models on humidity prediction task

Area	Evaluation metrics	CGC-BiLSTM	CGC	MMoE	Share Bottom	BiLSTM	CNNLSTM	LSTM	All-BiLSTM
Floor1	MAE	1.857	2.725	2.768	2.663	2.701	6.252	6.737	3.541
	CV-RMSE	0.062	0.085	0.082	0.087	0.093	0.176	0.181	0.099
	R ²	0.945	0.897	0.904	0.892	0.875	0.555	0.531	0.859
Floor2	MAE	1.640	2.240	2.510	2.403	2.706	5.678	6.222	2.490
	CV-RMSE	0.057	0.078	0.081	0.082	0.095	0.161	0.168	0.084
	R ²	0.955	0.913	0.907	0.906	0.873	0.633	0.602	0.900
Floor3	MAE	1.773	2.466	2.607	2.568	7.792	5.904	6.259	2.683
	CV-RMSE	0.069	0.092	0.091	0.095	0.226	0.192	0.194	0.099
	R ²	0.940	0.893	0.896	0.885	0.349	0.531	0.524	0.876
Mean	MAE	1.756	2.477	2.628	2.545	4.400	5.944	6.406	2.905
	CV-RMSE	0.062	0.085	0.085	0.088	0.138	0.177	0.181	0.094
	R ²	0.947	0.901	0.902	0.894	0.699	0.573	0.552	0.878

4.3 The impact of reducing the training set on indoor temperature prediction results

This study explores the performance effectiveness of the CGC-BiLSTM architecture in the task of indoor temperature prediction when the training set of environmental variables at a single point is insufficient. During testing, we observed the effect of model performance on all tasks when setting each of the three points as a target task and reducing the number of training sets. After successive tests, only the last 1500 datasets of the room temperature training set for the target task were eventually retained to simulate a training set insufficiency. The predictive power of the CGC-BiLSTM architecture for such tasks is explored by comparing the predictive effectiveness of all models in the target task and the predictive performance of the multi-task prediction model in other points prediction tasks.

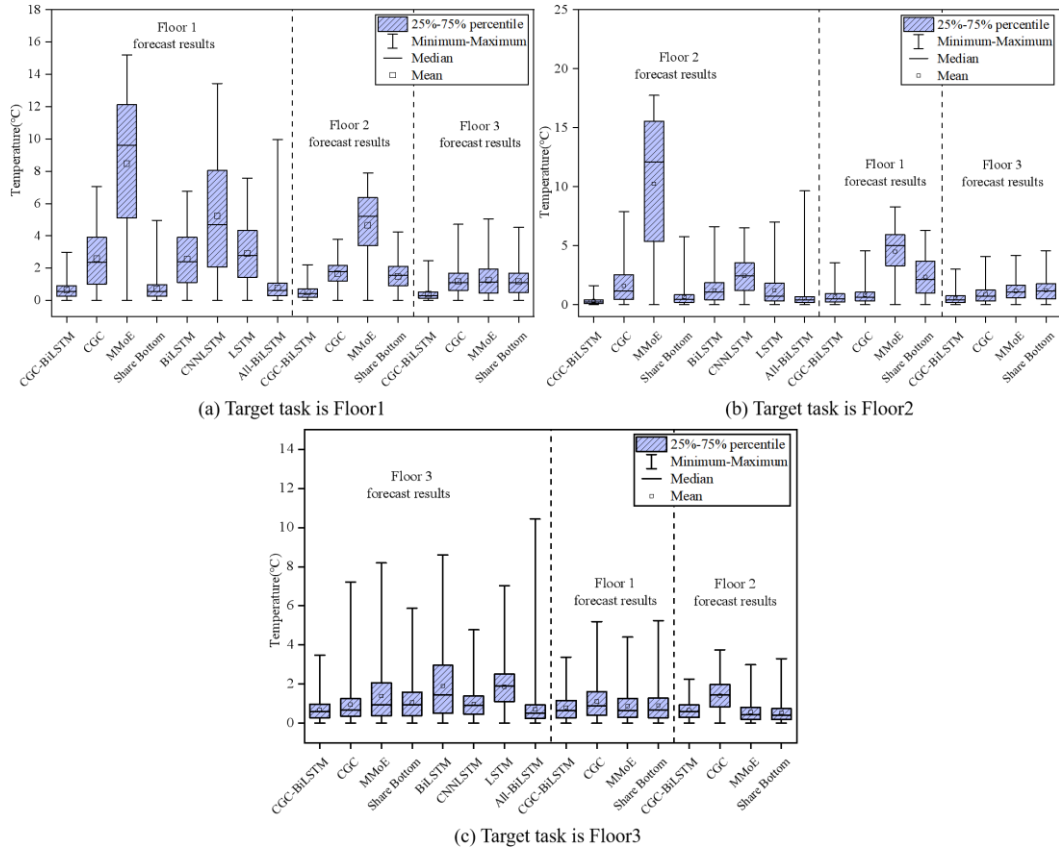


Fig.10. Error statistics of the temperature prediction task when setting different target tasks

Fig.10 show the prediction error statistics of all models on the temperature prediction task when different point is set as the target prediction task. On the target task since the training set is small so the prediction model produces large errors at certain time periods. From the overall prediction effect, CGC-BiLSTM is better than the other comparison models, and the error distribution is concentrated in a smaller error range. While comparing the results with all multi-task prediction models on the prediction tasks at other points, it was found that CGC-BiLSTM optimized the prediction results of the target task and had relatively small negative impacts on the other prediction tasks at the same time.

Table 4 describes the results of evaluation metrics calculated for different prediction models on the target task of indoor temperature prediction. Some of the single-task prediction models all performed poorly on the prediction task. Both multi-task prediction models, CGC and MMoE, also show poor prediction results, especially for CGC where the R^2 has been shown to be less than zero. It is shown that on the multi-point indoor temperature prediction task in this study, these two architectures may be less suitable for the situation of sparse training set for the target task set, and their complex structure leads to lower prediction results than the traditional Share Bottom architecture. Comparing the results of CGC-BiLSTM with all single-task prediction model evaluation metrics, we found that it is able to optimize the prediction effect in the case of a sparse training set of the target task through multi-task training architecture. On different target prediction tasks, CGC-BiLSTM can improve R^2 by an average of 6.44% and 5.31% compared to the optimal single-task and multi-task evaluation metrics computed.

Table 4

Calculation of evaluation indicators on the target task of indoor temperature prediction

Area	Evaluation metrics	CGC-BiLSTM	CGC	MMoE	Share Bottom	BiLSTM	CNNLSTM	LSTM	All-BiLSTM
Floor1	MAE	0.635	8.476	2.587	0.707	2.557	5.236	2.909	0.763
	CV-RMSE	0.045	0.536	0.178	0.053	0.172	0.360	0.192	0.062
	R^2	0.961	<0	0.390	0.946	0.425	<0	0.287	0.926
Floor2	MAE	0.304	10.241	1.613	0.637	1.218	2.433	1.229	0.534
	CV-RMSE	0.021	0.611	0.111	0.046	0.079	0.147	0.090	0.046
	R^2	0.984	<0	0.525	0.919	0.760	0.168	0.688	0.921
Floor3	MAE	0.676	1.387	0.949	1.060	1.888	0.975	1.847	0.704
	CV-RMSE	0.039	0.088	0.060	0.061	0.114	0.054	0.096	0.051
	R^2	0.933	0.660	0.843	0.835	0.426	0.871	0.593	0.886

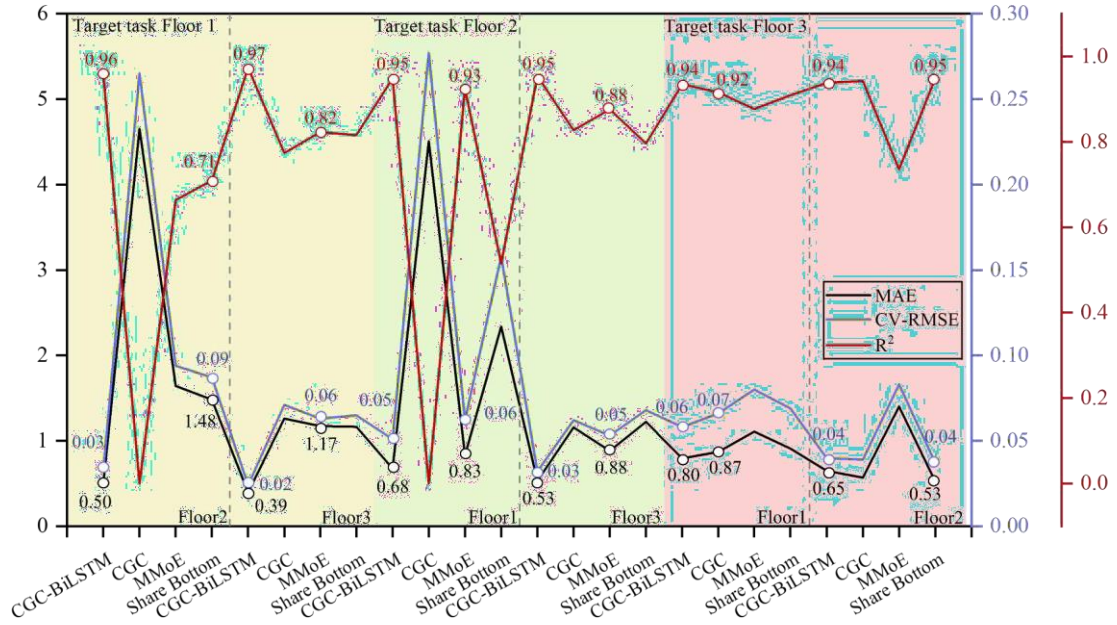


Fig.11. Calculation of evaluation metrics for the multi-task prediction model on the other temperature prediction tasks

As shown in Fig.11, the prediction results of all multi-task prediction models on the other tasks except the target task. We label the results of the evaluation metrics computed by CGC-BiLSTM on different tasks, as well as the results of other comparative models that R^2 is optimal in the same task. It can be found that in the experiments set up in this study, CGC-BiLSTM still performs well on the rest of the tasks except the target task, and the R^2 basically stays above 0.9. And the rest of the multi-task prediction models will show some decline in the prediction task for some experiments. Comparison of the evaluation metrics of CGC-BiLSTM for predicting the results of other tasks in both cases of complete and sparse training sets for the target task. R^2 decreased by an average of 2.14% and remained above 0.9. It is shown that the multi-task learning architecture used in this study optimizes the prediction effect on the target task while having less negative impact on other tasks in the multi-point indoor temperature prediction experiment.

4.4 The impact of reducing the training set on indoor humidity prediction results

The above study found that CGC-BiLSTM can show better prediction results when the training set of the target task is small in a multi-point indoor temperature prediction task. The experimental design process is as described in 3.2, and the humidity training set for the design target task is 3000 groups close to the test set due to the fast humidity changes with large fluctuations and other reasons that increase its prediction difficulty. The performance of CGC-BiLSTM in the multi-point indoor humidity prediction task is finally tested and its prediction results are compared with those of the single-task model and the multi-task model in the target task and other tasks.

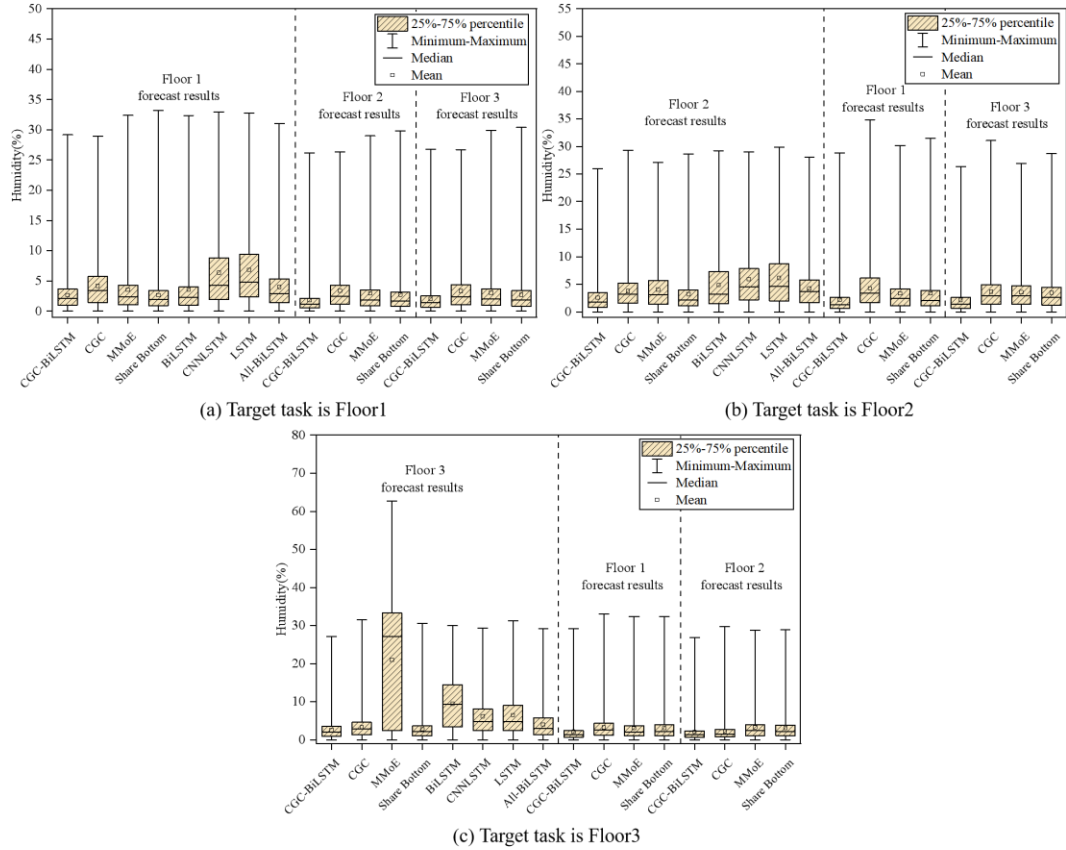


Fig.12. Error statistics of the humidity prediction task when setting different target tasks

As shown in Fig.12, the prediction results of single and multi-tasks on the target task and other tasks when the humidity prediction of different floors is set as the target task. It can be noticed that due to the more drastic fluctuations in humidity and the effect of the sparse training set, all models produce large prediction errors at some points. On the target task, CGC-BiLSTM can be better than the single-task prediction model by optimizing the performance effect of the model during the training process. Table 5 calculates the evaluation metrics for different model predictions on different target tasks. CGC-BiLSTM improved R^2 by 12.25%-16.60% compared to the best performing single-task model on different tasks. A comparison with other multi-task prediction models reveals that the multi-task architecture used in this study performs steadily on both the target task and other tasks. Compared with the optimal multi-task prediction model on different tasks, R^2 improves by 1.33%-7.09%.

Table 5

Calculation of evaluation indicators on the target task of indoor humidity prediction

Area	Evaluation	CGC-	CGC	MMo	Share	BiLST	CNNL	LSTM	All-
------	------------	------	-----	-----	-------	-------	------	------	------

	metrics	BiLST M		E	Botto m	M	STM		BiLST M
Floor1	MAE	2.684	3.557	4.193	2.683	3.536	6.365	6.814	3.999
	CV-RMSE	0.077	0.113	0.116	0.082	0.123	0.182	0.190	0.113
	R ²	0.916	0.816	0.808	0.904	0.784	0.522	0.481	0.816
Floor2	MAE	2.520	4.037	3.742	3.211	4.864	5.913	6.141	4.212
	CV-RMSE	0.075	0.114	0.099	0.100	0.141	0.164	0.173	0.113
	R ²	0.921	0.817	0.859	0.860	0.719	0.623	0.580	0.819
Floor3	MAE	2.604	21.00	3.298	2.778	9.418	6.149	6.454	4.035
	CV-RMSE	0.086	0.651	0.104	0.094	0.274	0.192	0.205	0.132
	R ²	0.906	<0	0.863	0.889	0.050	0.534	0.467	0.777

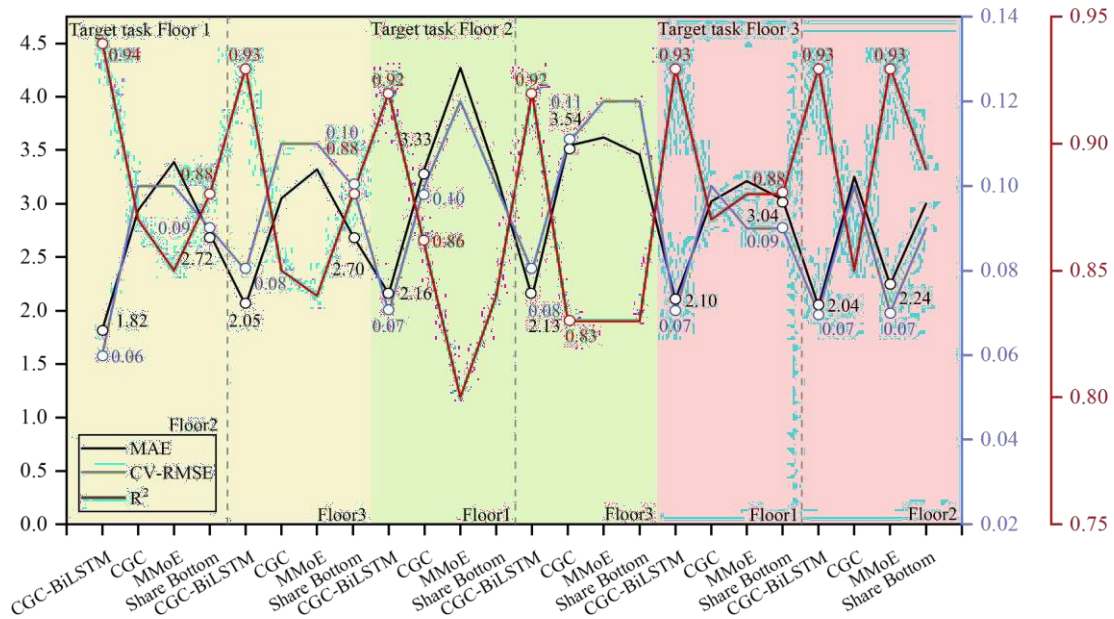


Fig.13. Calculation of evaluation metrics for the multi-task prediction model on the other humidity prediction tasks

Fig.13 describes the results of the evaluation metrics calculated for the multi-task prediction model on the other tasks when different target tasks are set. Taken together, CGC-BiLSTM is better than the comparison model on the other tasks. And the performance is smoother, with all R² above 0.9. CGC-BiLSTM on different tasks compared to other multi-task prediction models in the optimal evaluation metrics calculated, R² on average improved by 6% maximum 10.84%. The evaluation metrics for the remaining prediction tasks were calculated by comparing the two cases of complete and sparse training sets for the target task, and it is found that R² decreases by 1.92% on average. It is clear that on the dataset used in this study, CGC-BiLSTM can perform well on both the temperature and humidity prediction tasks.

5. Discussion

This study explores the performance of a multi-task prediction architecture combined with CGC and BiLSTM in predicting indoor environmental variables at multiple points in a connected open environment within the same office building. The research findings indicate that, under the complete

training set, the deep neural network incorporating the multi-task prediction architecture demonstrates better evaluation metric results for most tasks. This explains that under this architecture, simultaneously training multiple related points prediction tasks can optimize the overall prediction results to some extent. By reducing the amount of training data for temperature or humidity at different points and designating it as a target task, the study explores whether the model can improve the prediction performance of the target tasks while minimizing the negative impact on other tasks. The results show that most of the single-task prediction models used in this paper perform worse on the target task when a single point temperature or humidity prediction task is designed as the target task and the training set is reduced to 1500 and 3000 groups, respectively. In comparison, some of the multi-task prediction models show slight improvements over the single-task models in terms of prediction results for the target tasks, but they have affected the prediction accuracy of other tasks, resulting in a certain degree of negative impact. The architecture used in this paper separates the input parameters for different tasks and is designed with an expert network and a shared network layer, together with a residual network combined with an adaptive layer, which also improves the stability of the performance of the network in different tasks to some extent. Therefore, the insufficient training set for a single point has a minimal impact on the prediction performance of other tasks, resulting in a relatively superior comprehensive prediction outcome for multiple points, including the target tasks.

This study also faces certain limitations and shortcomings. Firstly, due to the constraints of the dataset, this study overlooked the impact of factors such as seasonality and external environmental changes when dividing the training and test sets. While this increased the difficulty of model training, it did not conduct separate studies for the cooling and heating seasons. Secondly, this study primarily focuses on the prediction tasks of indoor environmental variables at multiple related points within an open space, hence the test points were selected in corridors with a shared atrium. The performance of this approach in predicting tasks at multiple related points in large office buildings with other forms of cooling and heating systems still needs to be further explored. Due to the addition of a considerable number of network layers in the model architecture and the need to handle multiple tasks during training, the training time and difficulty of the model have been increased to some extent. Therefore, the architecture still needs further refinement to improve the processing capabilities for tasks with varying data set qualities, while also decreasing the training duration.

6. Conclusion

The main objective of this study is to construct an algorithmic architecture suitable for temperature or humidity prediction at multiple related points in the same building. When the training set at a point is insufficient, the architecture can still effectively accomplish the prediction task. For this purpose, we build multi-task predictive deep neural network algorithms combining neural networks such as BiLSTM under the CGC architecture. In order to verify the capability of the algorithm on this research task, we set up a series of comparison experiments on real datasets. The following conclusions are finally obtained.

The CGC-BiLSTM multi-task prediction model is compared with a single-task prediction model on the complete temperature or humidity dataset. The results show that the CGC-BiLSTM prediction results are all close to or better than the single-task prediction model. In different point temperature or humidity prediction tasks, the average R^2 is respectively improved by 3.51% and

4.99% compared with the optimal single-task prediction model. It is shown that this multi-task prediction algorithmic architecture can promote similar task learning in a multi-point temperature or humidity prediction task with less negative impact on different tasks.

The CGC-BiLSTM architecture is compared to other common multi-task algorithmic architectures on the complete temperature and humidity dataset. The algorithm was found to be overall better in different indoor temperature or humidity predictions. Among them, the average R^2 improvement on the humidity prediction task was significantly 7.86%. It is shown that the overall prediction is optimized to some extent by setting up a single expert layer and separating the input data for different tasks.

Finally, the results of different single-task and multi-task prediction models are compared by reducing the training set at a certain point and setting it as the target task. Most of the comparison models show a decrease in performance on the test set of the target task and in some cases are not able to predict data changes well anymore. In contrast, the CGC-BiLSTM model performs well on different temperature or humidity target prediction tasks. On the other tasks, CGC-BiLSTM is better than the other multi-task prediction models used for comparison in most cases, and the decrease in prediction results is smaller compared to the full dataset case. This suggests that the architecture optimizes the results of the goal prediction task while having less negative impact on other tasks.

However, there are many other cases of missing datasets that should be considered in real working conditions, such as missing data at random points or time periods. According to the actual situation of the building and the prediction task make appropriate adjustments to the algorithmic architecture and training methods, in order to obtain better prediction results.

Acknowledgment

The authors are grateful for the financial support of the National Natural Science Foundation of China (Grant No. 52078096 and 51806029), the 2023 Dalian University of Technology - Cardiff University Cooperation and Exchange Foundation Project, and 2023 International Exchange Foundation Project of "Co-Creation of Excellence Program" from Dalian University of Technology (Grant: DUTIO-ZG-202307), the Key Project of DUT for International Students Studying and Researching in China: Innovation and Practice of Talent Cultivation Model in the Field of Smart Buildings for the "Belt and Road" Initiative (Grant No. 1103-82120001) and the Natural Science Foundation Joint Fund of Liaoning Province (Grant No. 2023-MSBA-023).

References

- [1] Xu Y, Gao W, Qian F, et al. Potential analysis of the attention-based LSTM model in ultra-short-term forecasting of building HVAC energy consumption[J]. *Frontiers in Energy Research*, 2021,9:730640.
- [2] Wang Y, Shen G, Mehmani A. Indoor environmental wellness index (IEW-Index): Towards intelligent building systems automation and optimization[J]. *Building and Environment*, 2024,247:111039.
- [3] Mtibaa F, Nguyen K, Azam M, et al. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings[J]. *Neural Computing and Applications*, 2020,32(23):17569-17585.
- [4] Xu X, Zhong Z, Deng S, et al. A review on temperature and humidity control methods focusing on air-conditioning equipment and control algorithms applied in small-to-medium-sized buildings[J]. *Energy and Buildings*, 2018,162:163-176.
- [5] Li X, Han Z, Zhao T, et al. Online model for indoor temperature control based on building thermal process of air conditioning system[J]. *Journal of Building Engineering*, 2021,39:102270.
- [6] Wu H, Wong J W C. Temperature versus Relative Humidity: Which Is More Important for Indoor Mold Prevention?[J]. *Journal of Fungi*, 2022,8(7):696.
- [7] Wu H, Ng T W, Wong J W, et al. Environmental Sustainability and Mold Hygiene in Buildings[J]. *International Journal of Environmental Research and Public Health*, 2018,15(4):681.
- [8] Zhou X, Guo Q, Han J, et al. Real-time prediction of indoor humidity with limited sensors using cross-sample learning[J]. *Building and Environment*, 2022,215:108964.
- [9] Xiwen W, Xinli W, Xiaohong Y, et al. Distributed LSTM-GCN-Based Spatial–Temporal Indoor Temperature Prediction in Multizone Buildings[J]. *IEEE Transactions on Industrial Informatics*, 2024,20(1):482-491.
- [10] Xin S, Weiding L, Ying Z, et al. Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing[J]. *IEEE Access*, 2018,6:30559-30566.
- [11] Ahmad A S, Hassan M Y, Abdullah M P, et al. A review on applications of ANN and SVM for building electrical energy consumption forecasting[J]. *Renewable and Sustainable Energy Reviews*, 2014,33:102-109.
- [12] Ahmad M W, Mourshed M, Rezgui Y. Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption[J]. *Energy and Buildings*, 2017,147:77-89.
- [13] Talib A, Park S, Im P, et al. Grey-box and ANN-based building models for multistep-ahead prediction of indoor temperature to implement model predictive control[J]. *Engineering Applications of Artificial Intelligence*, 2023,126:107115.
- [14] Zamora-Martinez F, Romeu P, Botella-Rocamora P, et al. On-line learning of indoor temperature forecasting models towards energy efficiency[J]. *Energy and Buildings*, 2014,83:162-172.
- [15] Ming L, Yijun L, MinXinli. Practice and Application of LSTM in Temperature Prediction of HVAC System: 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)[C]. 12-14 June 2020.
- [16] Lin W, Zhang B, Li H, et al. Multi-step prediction of photovoltaic power based on two-stage decomposition and BiLSTM[J]. *Neurocomputing*, 2022,504:56-67.

- [17] Joseph L P, Deo R C, Prasad R, et al. Near real-time wind speed forecast model with bidirectional LSTM networks[J]. *Renewable Energy*, 2023,204:39-58.
- [18] Somu N, Sriram A, Kowli A, et al. A hybrid deep transfer learning strategy for thermal comfort prediction in buildings[J]. *Building and Environment*, 2021,204:108133.
- [19] Zang H, Liu L, Sun L, et al. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations[J]. *Renewable Energy*, 2020,160:26-41.
- [20] Meng A, Chen S, Ou Z, et al. A hybrid deep learning architecture for wind power prediction based on bi-attention mechanism and crisscross optimization[J]. *Energy*, 2022,238:121795.
- [21] Guixiang X, Jiancai S, Xiangfei K, et al. Prediction of Natural Gas Consumption for City-Level DHS Based on Attention GRU: A Case Study for a Northern Chinese City[J]. *IEEE Access*, 2019,7:130685-130699.
- [22] Yang Y, Gao P, Sun Z, et al. Multistep ahead prediction of temperature and humidity in solar greenhouse based on FAM-LSTM model[J]. *Computers and Electronics in Agriculture*, 2023,213:108261.
- [23] Fang Z, Crimier N, Scanu L, et al. Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model[J]. *Energy and Buildings*, 2021,245:111053.
- [24] Elmaz F, Eyckerman R, Casteels W, et al. CNN-LSTM architecture for predictive indoor temperature modeling[J]. *Building and Environment*, 2021,206:108327.
- [25] Ruiz C, Alaiz C M, Dorronsoro J R. A survey on kernel-based multi-task learning[J]. *Neurocomputing*, 2024,577:127255.
- [26] Misra I, Shrivastava A, Gupta A, et al. Cross-Stitch Networks for Multi-Task Learning: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C], 2016.
- [27] Ruder S, Bingel J, Augenstein I, et al. Sluice networks: Learning what to share between loosely related tasks[J]. *arxiv preprint*, 2018,arxiv:1705-8142.
- [28] Wu S, Peng D. Pre-SMATS: A multi-task learning based prediction model for small multi-stage seasonal time series[J]. *Expert Systems with Applications*, 2022,201:117121.
- [29] Qiao D, Wu S, Li G, et al. Wind speed forecasting using multi-site collaborative deep learning for complex terrain application in valleys[J]. *Renewable Energy*, 2022,189:231-244.
- [30] Liao Q, Cao D, Chen Z, et al. Probabilistic wind power forecasting for newly-built wind farms based on multi-task Gaussian process method[J]. *Renewable Energy*, 2023,217:119054.
- [31] Yongxin Y, Hospedales T. Deep multi-task representation learning: A tensor factorisation approach[J]. *arxiv preprint*, 2016,arxiv:1605-6391.
- [32] Jiaqi M, Zhe Z, Xinyang Y, et al. Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining[C], London, United Kingdom, 2018. Association for Computing Machinery.
- [33] Tang H, Liu J, Zhao M, et al. Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations: Proceedings of the 14th ACM Conference on Recommender Systems[C], Virtual Event, Brazil, 2020. Association for Computing Machinery.
- [34] Dingmei W, Ruixiao Z, Long Z. A Multivariate Load Prediction Method of Integrated Energy Systems Based on MMoE-TCN: 2022 7th International Conference on Power and Renewable Energy (ICPRE)[C].23-26 Sept. 2022.
- [35] Hye S J, Man H P, Man G N. Application of Multi-Task Learning for Abnormal Diagnosis and Trip Variable

Prediction in Nuclear Power Plants: 2022 International Conference on Computational Science and Computational Intelligence (CSCI)[C].14-16 Dec. 2022.

- [36] Tong J, Liu Z, Zhang Y, et al. Improved multi-gate mixture-of-experts framework for multi-step prediction of gas load[J]. Energy, 2023,282:128344.
- [37] Yan J, He Z, He S. Multitask learning of health state assessment and remaining useful life prediction for sensor-equipped machines[J]. Reliability Engineering & System Safety, 2023,234:109141.
- [38] Zhou K, Huang J, Han H, et al. Smart contracts vulnerability detection model based on adversarial multi-task learning[J]. Journal of Information Security and Applications, 2023,77:103555.
- [39] Chen Y, Xiao J, Wang Y, et al. Regional wind-photovoltaic combined power generation forecasting based on a novel multi-task learning framework and TPA-LSTM[J]. Energy Conversion and Management, 2023,297:117715.
- [40] Sima S, Neda T, Akbar S N. The Performance of LSTM and BiLSTM in Forecasting Time Series: 2019 IEEE International Conference on Big Data (Big Data)[C].9-12 Dec. 2019.
- [41] Baldi P, Brunak S, Frasconi P, et al. Exploiting the past and the future in protein secondary structure prediction [J]. BioinformaticsBioinformatics, 1999,15(11):937-946.
- [42] Houshy N, Giurgiu A, Jastrzebski S, et al. Parameter-Efficient Transfer Learning for NLP: Proceedings of the 36th International Conference on Machine Learning[C], Proceedings of Machine Learning Research, 2019. PMLR.
- [43] Liebel L, Körner M. Auxiliary tasks in multi-task learning[J]. arxiv preprint, 2018,arxiv:1805-6334.
- [44] Sameen M I, Pradhan B, Lee S. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment[J]. CATENA, 2020,186:104249.
- [45] Sameen M I, Pradhan B, Lee S. Self-Learning Random Forests Model for Mapping Groundwater Yield in Data-Scarce Areas[J]. Natural Resources Research, 2019,28(3):757-775.
- [46] Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for hyper-parameter optimization: Proceedings of the 24th International Conference on Neural Information Processing Systems[C], Granada, Spain, 2011. Curran Associates Inc.
- [47] Yu H, Zhong F, Du Y, et al. Short-term cooling and heating loads forecasting of building district energy system based on data-driven models[J]. Energy and Buildings, 2023,298:113513.