buildings

MDPI

Article

Revealing Microclimate around Buildings with Long-Term Monitoring through the Neural Network Algorithms

Xibin Wu [1], Jiani Hou [2], Jun Hui [1], Zheng Tang [1] and Wei Wang [3,*]

[1] State Grid Wuxi Power Supply Company, Liangxi Road 12#, Wuxi 210000, China; xibin_wu@163.com (X.W.); 13585085053@139.com (J.H.); griss_123@163.com (Z.T.)
[2] School of Computer Science and Engineering, Southeast University, Southeast University Road 2#, Jiangning District, Nanjing 211189, China; 213180375@seu.edu.cn
[3] School of Architecture, Southeast University, Sipailpou 2#, Xuanwu District, Nanjing 210096, China
* Correspondence: weiwang@seu.edu.cn

Abstract: The profile of urban microclimates is important in many engineering fields, such as occupant's thermal comfort and health, and other building engineering. To predict the profile of urban microclimate, this study applies the artificial neural network and long short-term memory network predictive models, and an urban microclimate dataset was obtained with a long-term monitoring from year 2017 to 2019 with 5-min resolution including temperature, relative humidity, and solar radiation. Two predictive models were applied, and the first (Model 1) is to apply the predictive techniques to predict the urban microclimate in the real-time sequence, and then extract the characteristics of urban microclimate, while the second (Model 2) is to directly extract the characteristics of the microclimate, and then predict the characteristics of the microclimate. Backpropagation artificial neural network (BP-ANN) and long-short term memory (LSTM) techniques were applied in both models. The results show Model 1 with as the time-series prediction can reach the best (99.92%) of correlation coefficient and 98% of the mean average percentage error (MAPE), for temperature, while 99.66% and 98.18% for relative humidity, respectively, while accuracies in Model 2 decreased to 79% and 88.6% of MAPE for temperature and relative humidity, respectively. The prediction of solar radiation using ANN and LSTM are 51.1% and 57.8% of the correlation coefficient, respectively.

Keywords: characteristics of urban microclimate; backpropagation artificial neural network; long-short term memory; long-term monitoring

check for updates

Citation: Wu, X.; Hou, J.; Hui, J.; Tang, Z.; Wang, W. Revealing Microclimate around Buildings with Long-Term Monitoring through the Neural Network Algorithms. *Buildings* **2022**, *12*, 395. https://10.3390/buildings12040395

Academic Editor: Geun Young Yun

Received: 22 January 2022
Accepted: 22 March 2022
Published: 23 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With urbanization accelerating in the world, especially in developing countries such as China, cities will host more than 60.6% of mainland China [1]. The urbanization also changes and shapes the urban microclimate semi-permanently [2], while urban microclimate plays an important role in creating urban sustainable environment and comfort of urban residents [3,4]. On the one hand, some studies proved the effect of urban climatic and environmental factors on heat-related illnesses, outdoor thermal stress [5,6], or building service and durability [7,8]. Some researchers have reviewed the different characteristics of urban microclimate, including solar radiant heat, ambient temperature, humidity, maximum and minimum daily temperature and humidity, and so on, and their impact on occupants' health [9,10], and occupants' perceptions [11,12]. Particularly, researchers found that weather condition (e.g., average temperature, and relative humidity) can influence COVID-19 transmission, and the most dangerous COVID-19 transmitter are PMs such as PM 2.5 and PM 10 [13]. On the other hand, urban microclimate, referring to the ambient air context around buildings rather than urban climate [14], has also become one of the most important issues influencing the production of buildings (such as their design, energy utilization, environmental conditioning, and so on) [15,16]. Researchers have proved through experimental and mathematical methods how ambient microclimate (temperature,

humidity, etc.) influences the potential of building energy systems and savings [17,18], regardless of whether it is an air conditioning system or a solar power system [19].

Knowing and predicting the characteristics and profile of an urban microclimate is an important prerequisite in the field of building science. For example, some researchers proved that ambient air temperatures forecasting is prevalent in the scientific community due to its economics and engineering applications, and some cases include the design of efficient solar energy production systems and sustainable buildings that depends on solar radiation and ambient temperature [20,21]. The profiles of urban microclimate (e.g., temperature, humidity, and solar radiation) have been studied in many works. Wu et al. discussed the impact different climate parameter on urban occupant' health with long-term urban climate data such as maximum, mean, and minimum temperature, humidity, and sunshine duration, and so on [22]. To predict profile of the urban microclimate, the most common methodology used is using mathematical modelling involving mass and energy balances or using machine learning techniques with the real-time series data from experiment. In the former, the one of microclimate (e.g., temperature) can be a formula of urban context. For example, Quemada-Villaomez proposed a simple mathematical model for estimating maximum and minimum daily environmental temperatures in a year [23]. However, this method, also called the white-box model, required good knowledge of urban climate and more related parameters in a formula. To overcome this, machine learning techniques, such as the black-box model, have become popular. The standard procedure for prediction is to feed the "machine" with a large amount of in situ data historical data.

Zhang et al. proposed long short term memory (LSTM) to predict urban micro-climate and further analyzed its impact on different-shape buildings [24]. Moghanlo et al. used an artificial neural network (ANN) model to predict the impact of climate changes in the Zanjan region, north-west Iran with daily meteorological data were analyzed from 2007 to 2018 and 1988 to 2018, respectively, and the observed climatic variables, including maximum and minimum temperature and precipitation, were determined as predictors in the artificial neural network [25]. Xie et al. predicted mean radiant temperature around buildings with ANN model [26]. Beroho et al. predicted the climate in northern Morocco successively with multilevel linear mixed effects for monthly temperature and rainfall observation between 1979 and 2014 and between 1984 and 2018, which shows a great prediction that can consider variance of each climate parameter [27]. Chang et al. proposed a model to develop the fine scale temperature prediction assisted with k-means clustering, support vectors machine, and independent dataset from the Community Weather Information Network (CoWIN) and the ASTER heat island intensity map [28]. Papantoniou and Kolokotsa used neural networks to predict outdoor air temperature and applied their predicitions in four European cities with a predictive horizon of 4–24 h, pointing out that with the collection of more outdoor data such as wind direction, precipitation rate and air pressure, neural networks will be developed for the prediction of wind speed and outdoor air humidity [29]. Bueno et al. used a computational model in urban weather generator and accounted for different urban morphologies and building usage distributions within a city to improve the model for an efficient temperature prediction to obtain an estimation of urban heat island [30]. With a six-month dataset, the study proposed the LSTM model to predict the maximum, average, and minimum values of temperature, air pressure, and clouds, demonstrating that LSTM is effective predictive technique for climate prediction [31].

Humidity is a sensitive parameter to occupant comfort model and building cooling energy use, and humidity limits have also been set up for quantify the thermal comfort [32]. Air humidity is also one important measurement parameter involved in many engineering and science fields, which can be stated in three different (absolute, specific and relative) ways, evaluating and predicting the fluctuation of which is increasingly necessary for a wide range of applications. Yang proposed a finite least-squares Fourier-series-based evaluation model to predict the daily 24-h air humidity ambient air humidity fluctuation. The model outperforms the two competitive models of autoregressive and ANN models [33]. Another parameter is solar radiation, which is not only an important determinant to building heating

gains but is also the key to renewable solar energy utilization. Bellido-Jim'enez generated and extracted the characteristics of urban climate (e.g., maximum and minimum daily air temperature) and integrated them into prediction of solar radiation using machine learning techniques such as ANN [34]. Guijo-Rubio et al. applied ANN model with three-type neural structures to obtain an extremely accurate prediction of the solar radiation from satellite images data [35]. Shboul et al. used ANN model for hourly solar radiation and wind speed prediction simultaneously [36], and a similar study was also conducted in [37].

Above all, besides the mathematical method, machine learning techniques, such as the neural network model, especially ANN and LSTM, have been the most popular predictive techniques in predicting urban microclimate. Current studies usually focus on the characteristics of temperature, such as maximum or minimum temperature, and rarely on the other characteristic of other urban microclimate, such as the cooling/heating degree days, maximum and average humidity and total solar radiation, respectively, even though the literature has stated their importance. Considering the difficulty of obtaining long-term and high-precision data through experiments, it would be more possible to have the limited granularity of data provided by urban weather stations, therefore, how effective the prediction of urban microclimate would be with different-granularity datasets. Therefore, with a long-term urban microclimate experiment from year 2017 to year 2019 with 5-min resolution including temperature, relative humidity, and solar radiation, this study selects the ANN predictive technique to generate and predict characteristics of urban microclimate through the ANN and LSTM techniques, especially. Two predictive models were applied, and the first is to apply the predictive techniques to predict the urban microclimate parameters of the real-time sequence, and then extract the characteristics of the microclimate parameters, while the second is to directly extract the characteristics of the microclimate, and then predict the characteristics of the microclimate. Therefore, this study can provide the insights how the predictive tools behave with long-term or limited dataset and contribute to the building engineering application that relies on urban microclimate.

2. Methodology

2.1. Overview of the Methodology

Figure 1 presents the overview of flowchart in the prediction process of microclimate, three parameters of which were selected, including temperature, relative humidity, and solar radiation, respectively. The testbed was the rooftop of one office building in Southeast University and microclimate was monitored through local or cloud storage. In the data pre-processing, the resolution of data was reduced from 5 min to 1 h to release the computation burden, and normalization step and data characteristics are necessarily by abstracting the features of time-series microclimate data. Two models were created to predict the characteristics of time-series microclimate data: One uses real-time historical microclimate data with predictive techniques to predict the future real-time microclimate, then abstracting characteristics of microclimate, name (model 1); The second abstracts characteristics of microclimate first to consist of new time-series data, then uses predictive techniques to predict future characteristics of microclimate. Those characteristics of microclimate include the maximum, average, and minimum temperature, and humidity, cooling and heating degree days, and total amount of solar radiation in a day.

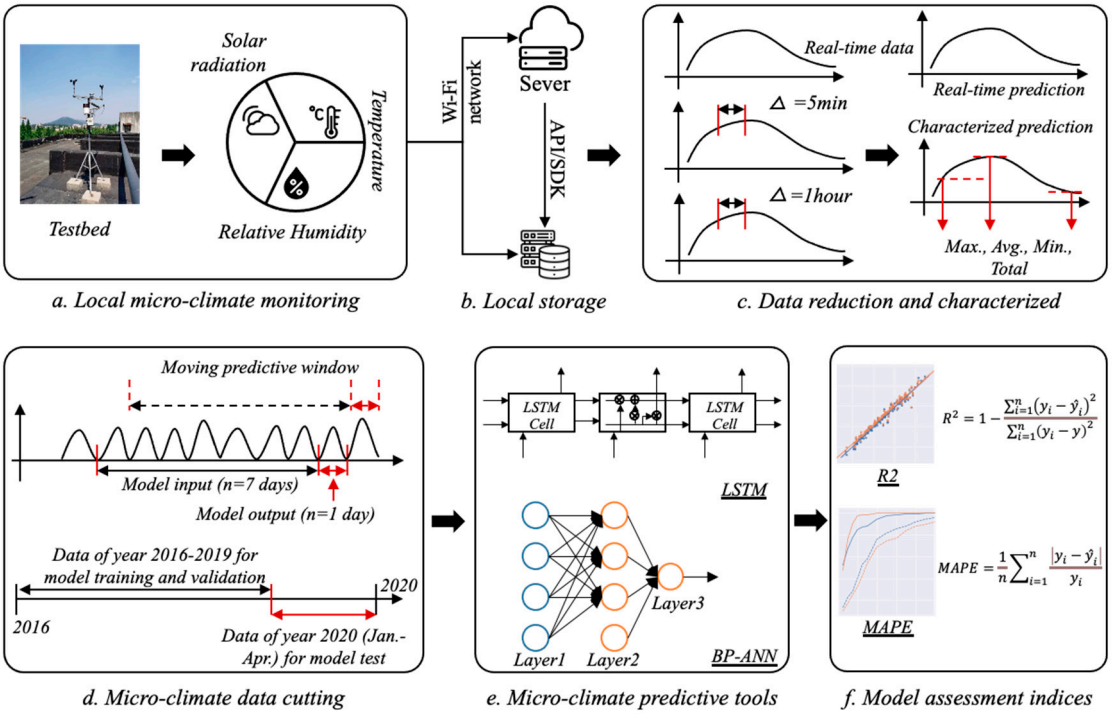


Figure 1. The overview of the methodology for microclimate prediction in this study.

2.2. Case Experiment of Urban Microclimate Monitoring

In this experiment, the microclimate station was installed on the roof of a seven-floor building (Log: 118.80, Lat: 32.06), which is one of the buildings meeting the two requirements in Southeast University, Xuanwu District, Nanjing, Jiangsu Province, as presented in Figure 2. The microclimate station was from the Onset Computer Corporation. The data can be stored locally as well as online server through Wi-Fi channel. The parameters used in this study consist of the temperature (°C), relative humidity (%), rain (mm), solar radiation (W/m^2). The case experiment started from January 2017 to April 2020 with the resolution of 5 min. Table 1 concludes the basic information about version, operating temperature, accuracy, resolution, measurement range and size of the selected sensors in this study.

Table 1. The basic information of sensor set used in this study.

	Data Logger	Temperature(T)/Relative Humidity (RH)	Solar Radiation
Version	RX3003	S-THB-M002	S-LIB-M003
operating temperature	40–60 °C	−40–75 °C	−40–75 °C
Accuracy	-	T: ±0.21 °C RH: ±2.5%	±10 W/m^2 or ±5%
Resolution	-	T: 0.02 °C RH: 0.1%	1.25 W/m^2
Measurement range	-	T: −40–75 °C RH: 0–100%	0–1280 W/m^2
Size	186 mm (H) × 181 mm (L) × 118 mm (W)	10 × 35 mm	41 mm (H) × 32 mm (Φ)

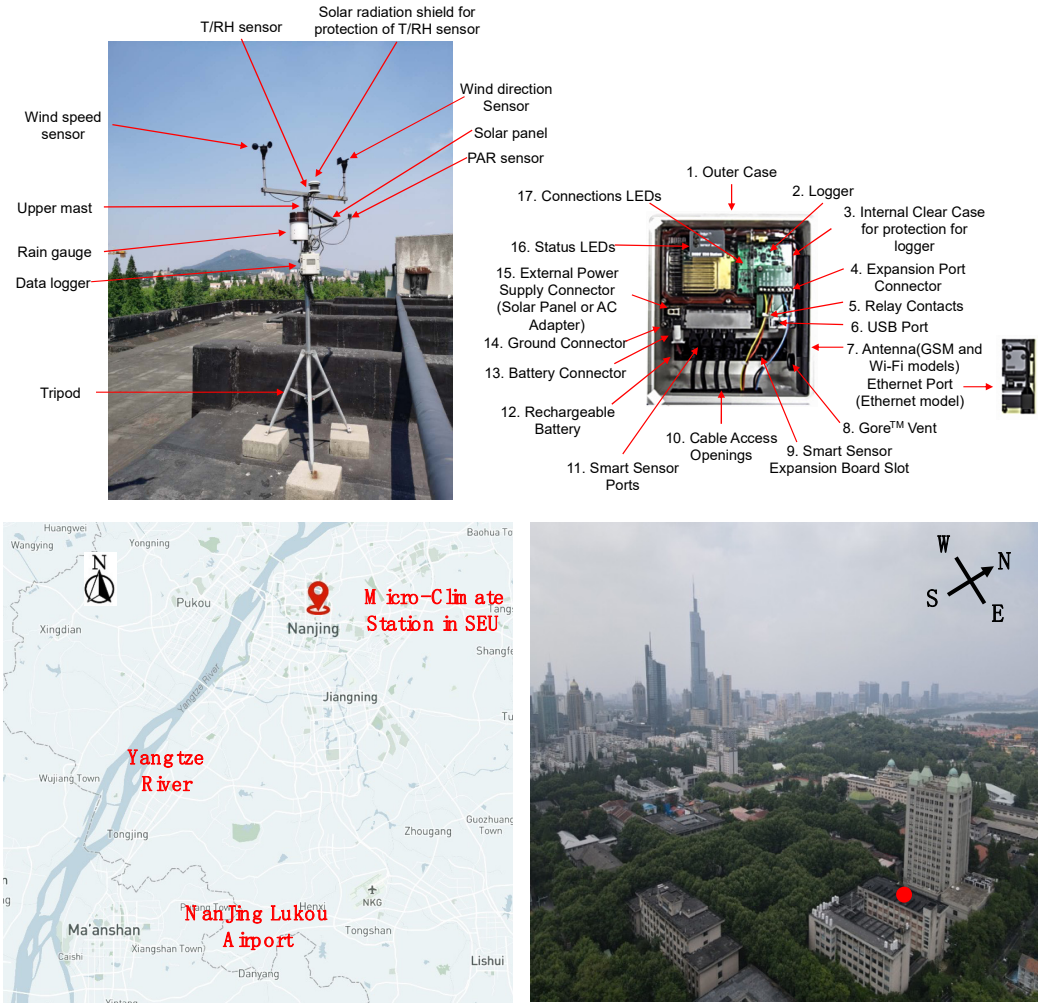


Figure 2. The picture and location of micro-climate station in this study.

2.3. Data Pre-Processing

Before performing predictive techniques, the data normalization is required and this study selects the zero-mean normalization method, as presented in Equations (1) and (2), to transform the data within the range of -1 to 1. In the equations, x is the original data and x' is new transformed data. μ and δ are the mean and standard deviation of original data.

$$x' = \frac{x - \mu}{\max(x) - \min(x)} \tag{1}$$

$$x' = \frac{x - \mu}{\delta} \tag{2}$$

For characteristics of microclimate, this study abstracts their maximum, average, and minimum values respectively of temperature and relative humidity. While for solar radiation, the total amount of solar radiation was calculated by summing the values within a range of period. In those equations, the abstraction of relative humidity is the same as it of temperature and n is set as 23 as one day. Degree days are based on the assumption that when the outside temperature is reference temperature (e.g., 18 °C or 65 °F), building don't need heating or cooling to maintain the comfortable [38]. The Cooling Degree Days (CDD) and Heating Degree Days (HDD) can be calculated by Equations (3) and (4). The most

common use of degree days is for tracking energy use. A CDD or HDD is a measurement designed to quantify the demand for energy needed to cool or heat buildings.

$$HDD_i = \frac{1}{n}\int (T_{BP} - T_{Avg.})dt \text{ for } T_{BP} > T_{Avg.} \quad (3)$$

$$CDD_i = \frac{1}{n}\int (T_{Avg.} - T_{BP})dt \text{ for } T_{BP} < T_{Avg.} \quad (4)$$

where HDD_i is heating degree days for one day (°C), CDD_i is cooling degree days for one day (°C), T_{BP} is the balance point temperature. $T_{Avg.}$ is the daily average temperature (°C) calculated from the real-time weather data.

2.4. Predictive Techniques in This Study

The predictive techniques consist of typical backpropagation artificial neural network (BP-ANN) and long short-term memory (LSTM) models, which are typically applied in the time-series prediction. Those two models were performed in Python 3 with PyCharm interpreter.

2.4.1. BP-ANN Predictive Technique

BP-ANN is a fully connected neural network, the structure of which is shown in the Figure 1. The structure includes a three-layer network, which is an input layer, a hidden layer, and an output layer. The input layer inputs known data, the hidden layer performs nonlinear calculations to extract data features, and the output layer outputs the required data., which only has one. In this experiment, model 1 has input eight data (seven consecutive days of meteorological parameter characteristic value data, and a mean value) and output one data, namely the predicted value of the eighth weather phenomenon parameter characteristic value. Model 2 has input twenty-five data (24-day hourly weather parameter data, and an average value) and output one data (the predicted value of the weather parameter selected at the next moment).

2.4.2. LSTM Predictive Technique

LSTM is a special neural network, recurrent neural network (RNN), mainly used for training long sequences. LSTM contains three gates, namely a forget gate, input gate, and output gate. The forget gate is used to determine the information that needs to be discarded when the cell state, the input gate is used to determine the information that needs to be updated in the current cell state, and the output gate is used to determine the current cell state. The output information of the cell state, the processing information of these three gates is summarized, and a final output of the LSTM cell state is obtained. Their working principle can be summarized as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$f_t = \sigma\Big(W_f[h_{t-1}, x_t] + b_f\Big) \quad (6)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (7)$$

$$\widetilde{C}_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (8)$$

$$C_t = f_t * C_t + i_t * \widetilde{C}_t \quad (9)$$

$$h_t = o_t * \tanh C_t \quad (10)$$

where $*$ represents matrix multiplication, i_t, f_t, o_t are the outputs of different gates, $\widetilde{C}_t$ is the candidate new state of the current cell, C_t is the new cell state, h_t is the final output of the current cell, W_i, W_f, W_o are weight matrices, b_o, b_f, b_i are deviation matrices. Through the role of different gates, LSTM cells can obtain complex interrelated information between long sequences.

2.5. Model Tunning of Two Predictive Techniques

Two scenarios were defined in this prediction process, which is presented in Figure 3. This study mainly proposes an approach of predicting characteristics of urban microclimate in two scenarios. While validating the approach with two scenarios, one long-term dataset from urban microclimate monitoring was applied and the two prediction methods are respectively extracting the feature values of the meteorological parameter data from January 2017 to April 2020, and then using the obtained feature value data from January 2017 to December 2019 as the training dataset to predict the characteristic values of the meteorological parameters from January to April 2020, which are the test dataset. After the characterization, the dimensionality of those datasets was reduced from 5 min to 1 h and 1 day. Model 1 uses 24 h as the step size plus the average difference to predict the next (h) data. Model 2 The training set uses seven (day) as the step size plus the average difference to predict the next (day) data. Average difference calculation can be described as:

$$\delta = \frac{1}{k}(x_n - x_{n-1} + x_n - x_{n-2} \ldots x_n - x_{n-k}) \tag{11}$$

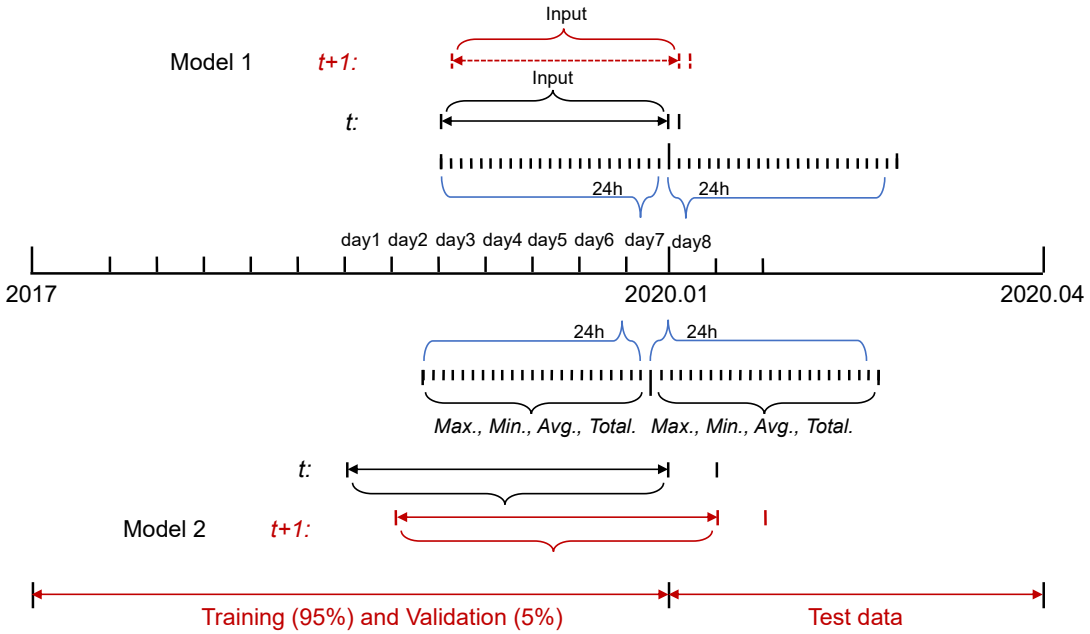


Figure 3. Illustration of data generation process in two predictive models using BP-ANN and LSTM.

Parameter tunning is the necessary step for predictive process. The mean average percentage error (MAPE) and correlation coefficient were applied in the model tunning process for the gradient loss learning. To find the suitable technique, each kind of predictive model was adjusted. The sizes of original and reduced urban microclimate training datasets are 315,360 (5-min resolution) and 26,280 (1-h resolution) for real-time prediction and 1097 for characteristic prediction. The range of batch size was 16, 32, 64, 128; the epochs are from 60 to 180 in the step of 10, and optimizer is "leakyRele", "Adam" and "SGD". In model 2, the range of batch size was 16, 32, 64, 128; the epochs are from 60 to 180 in the step of 10, and optimizer is "leakyRele", "Adam" and "SGD". While for the validation test, the correlation, mean absolute percentage error and x-accuracy are applied, respectively. The equations are as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \tag{12}$$

$$r(\hat{y}, y) = \frac{Cov(\hat{y}, y)}{\sigma_{\hat{y}}\sigma_y} \tag{13}$$

$$\tau(\hat{y}_i, x) = \frac{\sum_1^M X(|\hat{y}_i - y_i|,\ x)}{n}$$
$$X(|\hat{y}_i - y_i|,\ x) = \begin{cases} 1, & |\hat{y}_i - y_i| \leq x \\ 0, & |\hat{y}_i - y_i| > x \end{cases} \quad (14)$$

where, $\hat{y}_i$ and y_i are the predicted value and ground truth, respectively. $r(\hat{y}, y)$ is the correlation coefficient and $\sigma_{\hat{y}}, \sigma_y$ are the standard deviation. $\tau(\hat{y}_i, x)$ is the x-accuracy of a prediction model when the allowable error is smaller than x, n is the sampling size, and x is the tolerance level.

3. Results

3.1. Results of Long-Term Microclimate Monitoring

Figure A1 presents the example of distribution of temperature, relative humidity, and solar radiation, respectively, for typical days (1st, January and 1st July) for three years. The figure shows that in the typical days, the temperature decreases by years, combined with the Figures A2 and A3, which present the maximum and minimum temperature distribution, its clearly to find that the temperature could increase by months in the years, which says that the maximum and minimum temperature in year 2017 decrease from January to December compared with them in year 2019. This could conclude that temperature has a cooler trend in summer and a warmer trend in winter. Comparing the relative humidity in three years, Figures A4 and A5 clearly show that relative humidity is much more random than temperature, having no clear trend. The relative humidity is also influenced by several factors, such as rain, temperature, and so on. The randomness is also found in Figure A1 for two single days and the characteristics can also bring challenges of prediction accuracy. As for solar radiation in Figure A6, it only appears during the day and peaks at about noon, forming a discrete type of data. In terms of seasons, the total solar radiation is mostly gradually increasing, reaching a peak in summer and a valley in winter. Comparing the distributions in three years, the total amount of solar radiation shows a downward trend every month. It is very interesting to find that the solar radiation is decreasing, although the temperature is increasing.

3.2. Results of Predictive Models

This subsection discusses the prediction results of temperature characteristics from Jan. to Apr. in year 2020 with long-term monitoring from year 2017 to 2019. In the parameter tuning, for BP-ANN, batch_size is 64, epoch is 120, and optimizer is Adam, while for LSTM, batch_size is 64, epochs is 100 and optimizer is Adam. Figures 4–6 show the comparative results from BP-ANN and LSTM to actual value. The results clearly show that although the two models can achieve significantly accurate prediction of maximum, average, and minimum temperature value, the real-time predictions in model 1 have more accurate prediction than characteristic predictions in model 2 do, no matter inferred from the fitting between different curves and correlation coefficient plotting. The reason why real-time prediction is more accurate is due to a sufficient amount of data to feed the prediction model, for example, 157,680 pieces of data in three years are used for real-time prediction in model 1 while only 1095 pieces of data in three years for characteristic prediction in model 2.

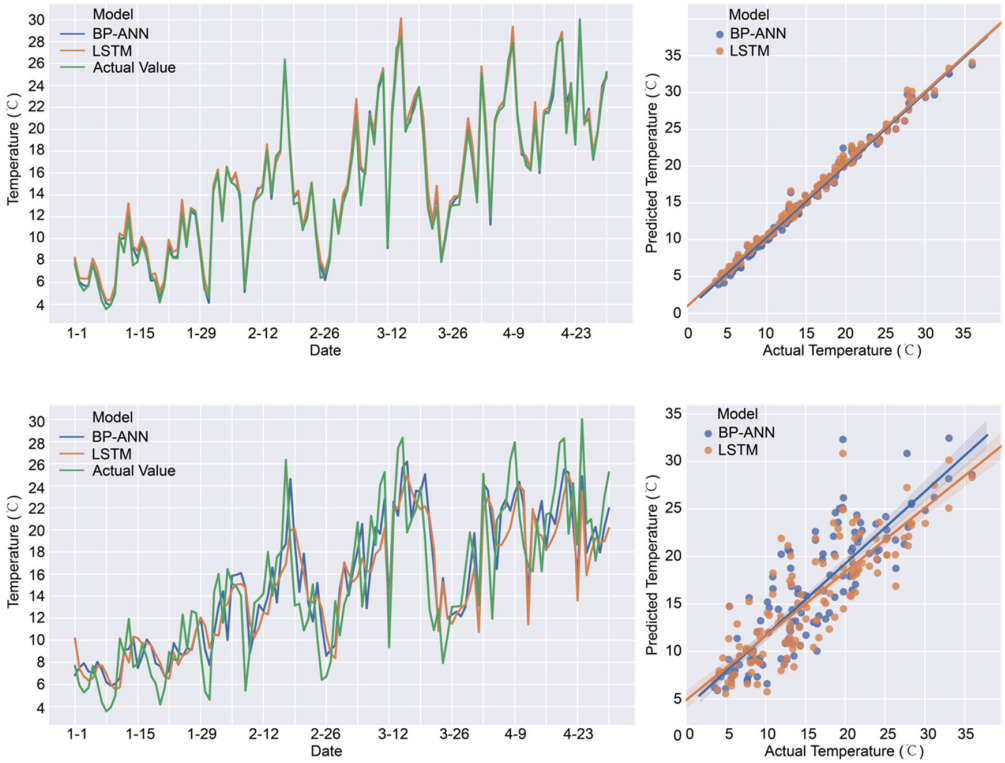


Figure 4. The prediction results of daily maximum temperature using BP-ANN and LSTM techniques in Model 1 (**up**) and 2 (**below**).

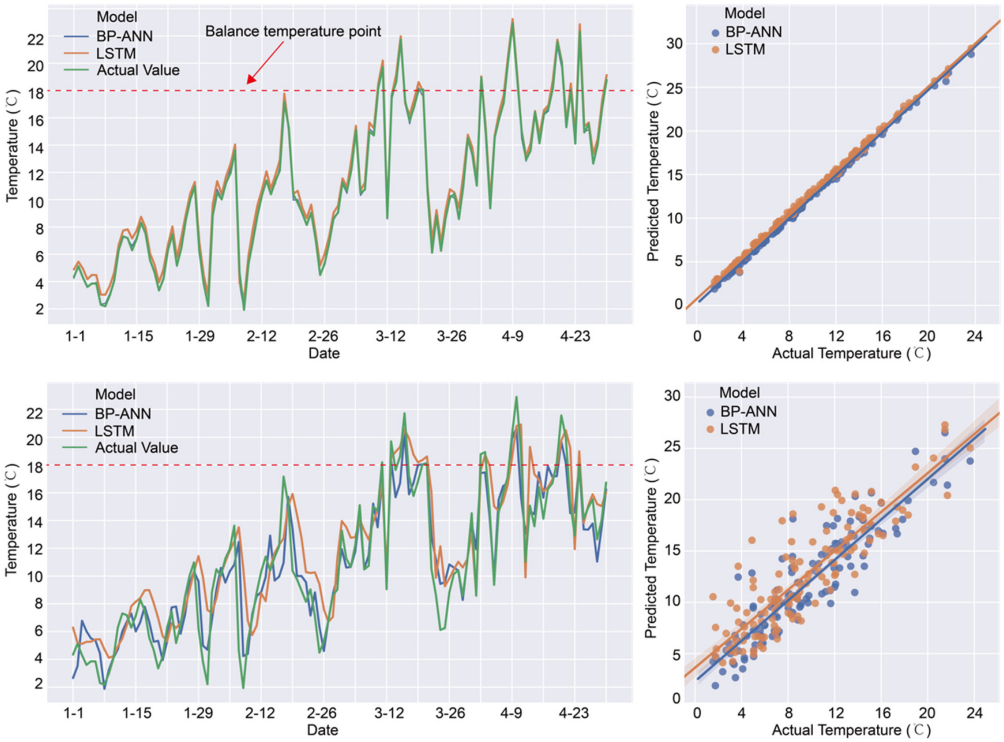


Figure 5. The prediction results of daily average temperature using BP-ANN and LSTM techniques in Model 1 (**up**) and 2 (**below**).

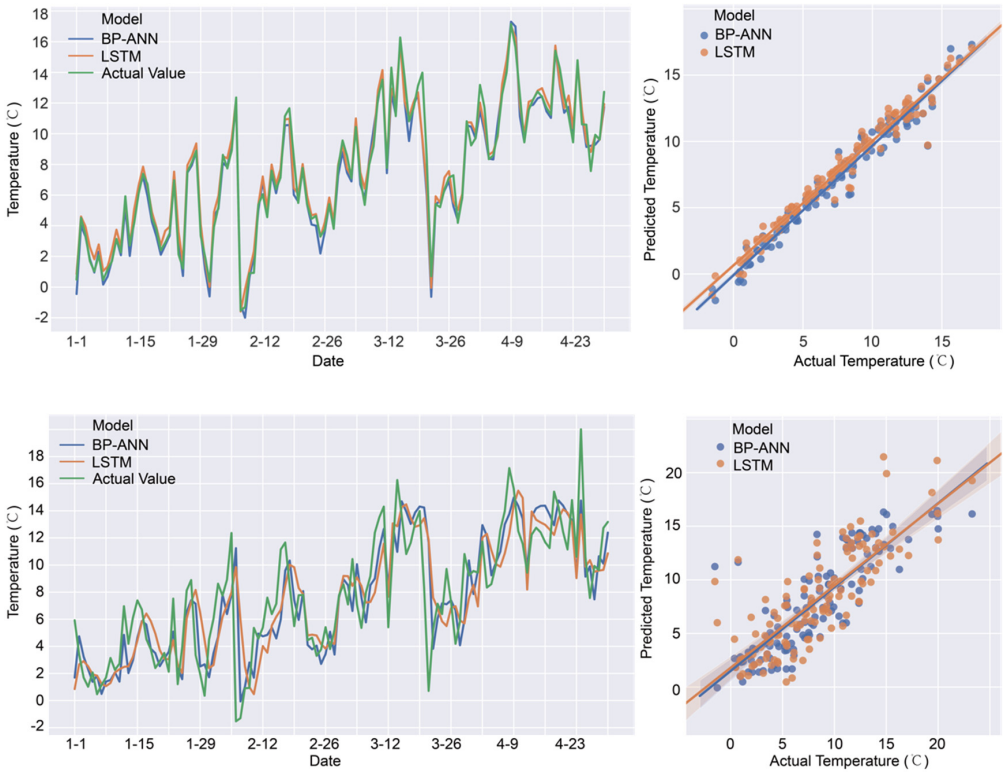


Figure 6. The prediction results of daily minimum temperature using BP-ANN and LSTM techniques in Model 1 (**up**) and 2 (**below**).

Additionally, Table 2 and Figure 7 summarize the numerical prediction accuracies, mainly comparing the results of correlation coefficient and MAPE, and x-accuracy. From the results, it can be seen that the real-time prediction accuracy in model 1 is the highest, especially when comparing correlation coefficient. In addition, when comparing correlation coefficient and MAPE, it can be found that the accuracy of the average temperature prediction is the highest, up to 98% and 93% accuracy, respectively using BP-ANN and LSTM, while the minimum temperature prediction is the worst. At the same time, in comparison, the accuracy of BP-ANN is better than that of LSTM, although both can obtain highly acceptable results. Compared with the eigenvalue prediction model, the prediction accuracy is greatly reduced. From the perspective of correlation coefficient, although it is also in the acceptable range, from the perspective of MAPE, the best accuracy is only 79%. Looking into x-accuracy results, in the characteristic prediction of temperature in model 2, for the maximum temperature, the allowable error is within 4 °C and the overall prediction accuracy of BP-ANN technique can reach 80% or more. As for the average temperature, only the allowable error is Within 3 °C, and the same is true in the prediction of the minimum temperature. In summary, the prediction will suit for average temperature prediction, then maximum temperature.

Table 2. The correlation coefficient and MAPE assessment indices for daily temperature prediction results of two models using BP-ANN and LSTM techniques.

		Correlation Coefficient			MAPE		
		Max.	**Avg.**	**Min.**	**Max.**	**Avg.**	**Min.**
Real-time prediction	BP-ANN	99.52%	99.93%	98.62%	0.04	0.02	0.14
	LSTM	99.51%	99.92%	98.72%	0.06	0.07	0.13
Characteristic prediction	BP-ANN	87.06%	91.07%	86.56%	0.24	0.21	0.47
	LSTM	84.47%	86.49%	82.51%	0.26	0.32	0.51

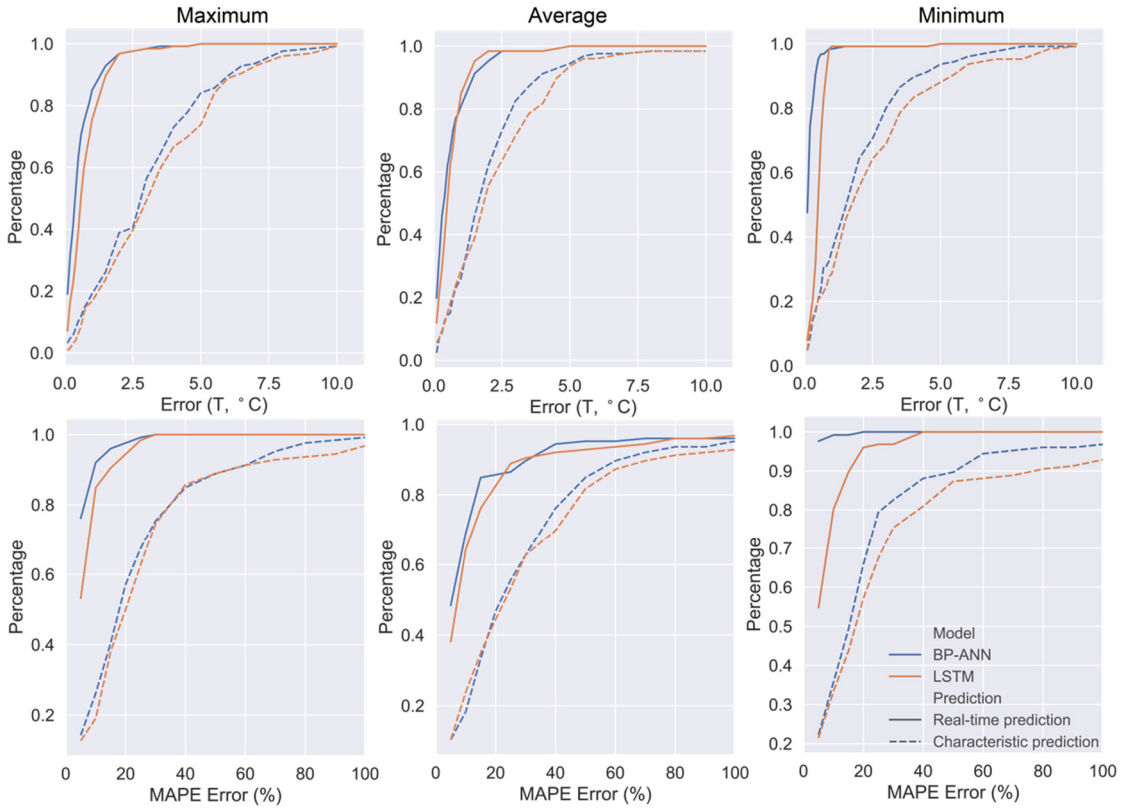


Figure 7. The x-accuracy prediction results of daily temperature using BP-ANN and LSTM techniques in Model 1 and 2.

Figure 8 concludes the comparison of HDD results between BP-ANN and LSTM predictive techniques in model 1 and 2. HDD is a parameter that implies the heating load of a building in winter. It is a parameter closely related to the energy consumption of the building. It can be seen from Equations (7) and (8) that this parameter is highly correlated with the average temperature. It can be seen from the histogram results that in the two prediction models (model 1 and 2), there is no obvious difference, except for model 1, the HDD results predicted by LSTM are lower than other models. The reason for this result may be that each model's prediction of the average temperature is more accurate. This also shows that the results of using the two models to predict the temperature characteristics will not have a big impact on the estimation of energy consumption, and will not cause excessive differences when sizing energy-related equipment.

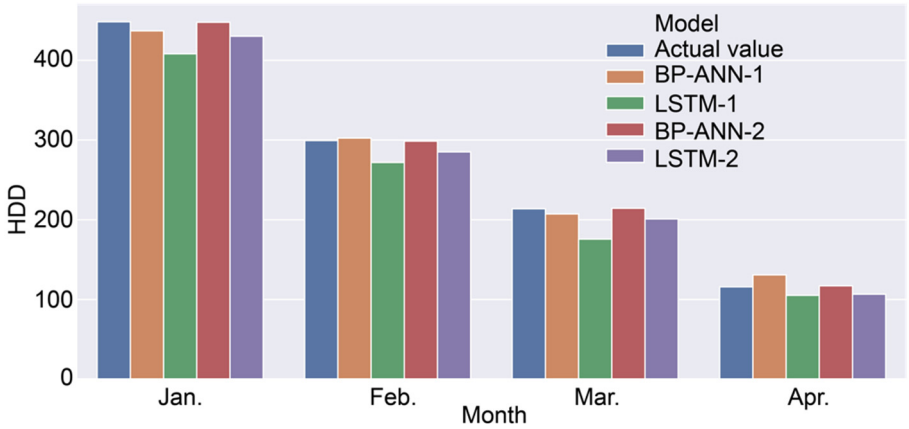


Figure 8. The comparison between HDD results using BP-ANN and LSTM techniques in Model 1 and 2.

Figures 9–11 illustrate the prediction results of RH parameter in the microclimate and similar conclusion can be drawn to the prediction result of temperature, so for brevity, the details of results omitted. Similarly, real-time prediction is much more accurate than characteristic prediction. The same reason may be that real-time prediction provides more available datasets. Also, the prediction of the maximum relative humidity value is not as accurate as the average humidity prediction. It can also be seen from the existing results that LSTM is more suitable for the prediction of relative humidity.

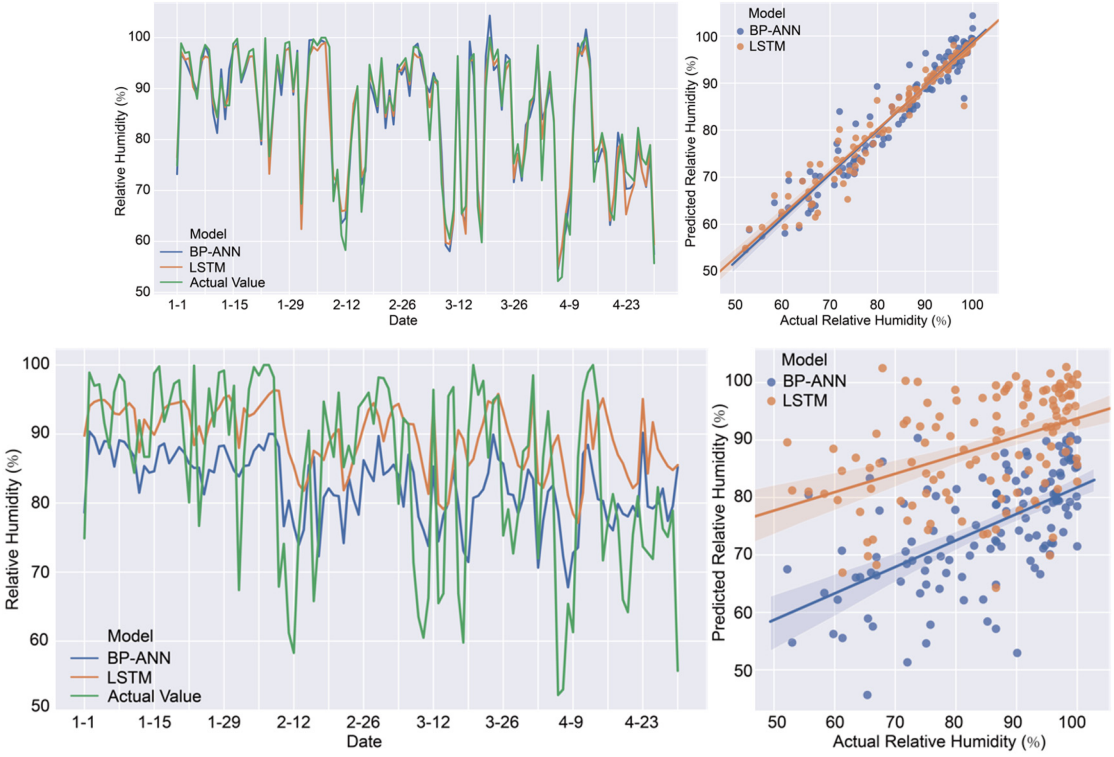


Figure 9. The prediction results of daily maximum RH using BP-ANN and LSTM techniques in Model 1 (**up**) and 2 (**below**).

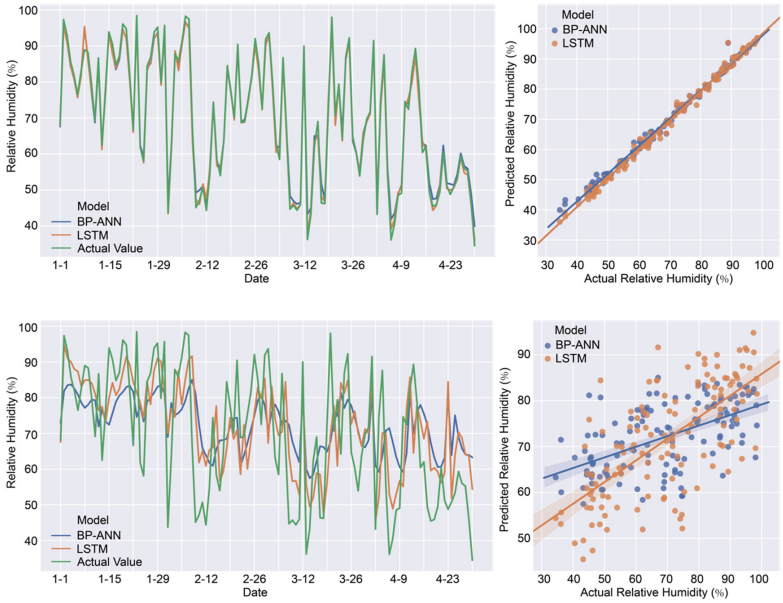


Figure 10. The prediction results of daily average RH using BP-ANN and LSTM techniques in Model 1 (**up**) and 2 (**below**).

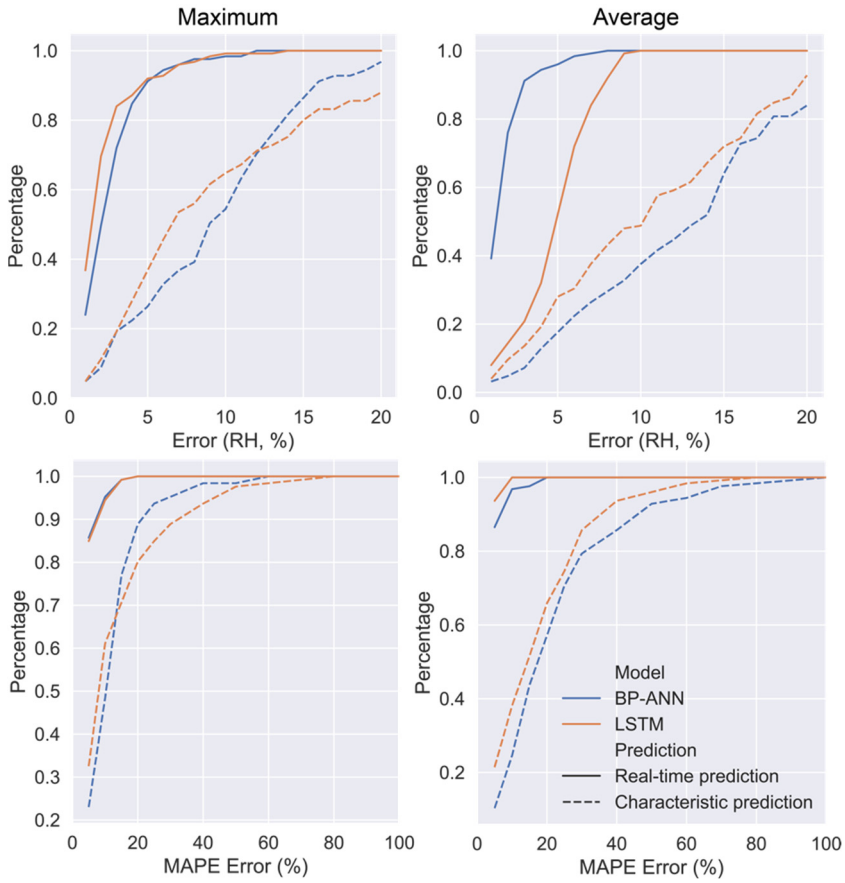


Figure 11. The x-accuracy prediction results of daily RH using BP-ANN and LSTM techniques in Model 1 and 2.

Inferred from Table 3 about the correlation coefficient and MAPE results, correlation coefficient shows great accuracies of prediction results in real-time and characteristic RH results, among which the best correlation coefficient can achieve 99.66%. The same poor performance can be found for prediction of characteristics of RH in Table 3. Comparatively, RH is more volatile and less stable than temperature, and when the characteristic (maximum and mean) of RH were predicted through characteristics, prediction accuracies are poor in terms of trend no matter BP-ANN and LSTM, although the accuracies in terms of relative error (MAPE) is acceptable. It also concludes that real-time dataset is much better for RH prediction. As for results of solar radiation in Figure 12, no matter the prediction fitting curves or the x-accuracy, the result clearly shows a relatively poor prediction, whether it is BP-ANN or LSTM. It can be seen from the results that the correlation of the results is poor, which of BP-ANN and LSTM are 51.1% and 57.8%, respectively. The reason for this might be that this study only considers the prediction of total daily solar radiation, which is also a more useful value according to the attribute of solar radiation used for the evaluation of solar energy utilization potential. But due to this, the granularity of the data is reduced, and the data used for training is relatively reduced, another reason might be that the fluctuation of the data itself is also relatively large, which has a greater impact on the prediction results. Therefore, on the one hand, the prediction of solar radiation requires further optimization of the algorithm itself or further mining of the law of the data itself, to predict the long-term solar radiation more effectively, in turn to more accurately predict building energy use and solar energy utilization. On the other hand, the fluctuation of solar radiation is also related to the weather of the day, especially rainy or cloudy days. Therefore, to accurately predict solar radiation, it might be necessary to introduce other weather parameters to data-joint prediction, which needs further discussion, however, is ignored in this study for the sake of brevity.

Table 3. The correlation coefficient and MAPE assessment indices for daily RH prediction results of two models using BP-ANN and LSTM techniques.

		Correlation Coefficient		MAPE	
		Max.	**Avg.**	**Max.**	**Avg.**
Real-time prediction	BP-ANN	96.78%	99.66%	3.09%	2.63%
	LSTM	97.48%	99.66%	2.59%	1.92%
Characteristic prediction	BP-ANN	59.00%	55.65%	11.40%	21.54%
	LSTM	43.57%	68.23%	12.79%	17.28%

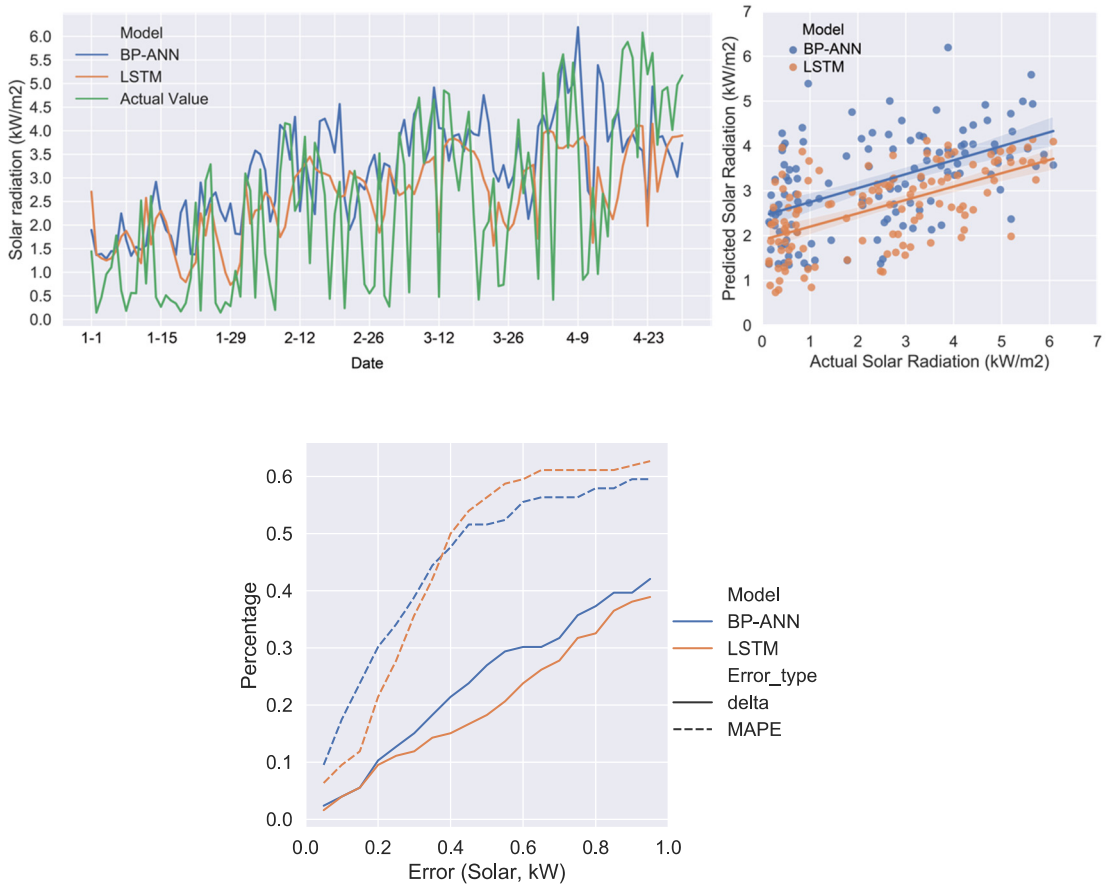


Figure 12. The prediction of total solar radiation and its x-accuracy results BP-ANN and LSTM techniques.

4. Discussion

The profiles and characteristics of urban microclimate are important as discussed in literatures. However, such kinds of predictions with machine learning techniques were rarely applied for different characteristics of different urban microclimate parameters. Meanwhile, such predictions are also important in characterizing urban microclimate with long-term or incomplete datasets, the latter of which could be the habitus. Although this study discussed the application of Model 1 to predict the characteristics of urban microclimate with long-term monitoring and model 1 also shows a good prediction accuracy, it is usually difficult to acquire long-term urban microclimate dataset, therefore limiting its wide application. In the Model 2, only the data of characteristics of urban microclimate was obtained, although the accuracies are lower, such dataset are usually easier to acquire as open data from most of climate station or official climate report. Therefore, Model 2 could be the options of predictive model to illustrate the profiles of microclimate. Those are why this study proposed two prediction models.

With the benefits of those predictive models, several building engineering can be realized, for example, (i) to determine the thermal comfort of urban environment from prediction result; (ii) estimating and predicting cooling and heating load with average temperature using CDD and HDD; (iii) the characteristics of solar radiation (maximum and total) are quite important for roof-top solar renewable energy utilization. However, in such studies, even dataset of characteristics of urban microclimate is much more easily to acquire, more data is still needed to feed the predictive algorithm and also more efforts are required to train the algorithms to receive accurate predictions, especially for solar radiation.

This study also contains some limitations. The first one is the fact that although there are several building engineering applications above, this study (for brevity) did not put forward to validate and test for the future engineers, which could be the next works. Secondly, this study conducted one long-term microclimate monitoring experiment, one feature of microclimate was not considered in this study, however important for understanding microclimate, which is its periodic variation. Meanwhile, urban morphology should be also an important supplementary parameter to illustrate urban microclimate. Therefore, future works should pay more efforts on those points.

5. Conclusions

This study conducted the work that generates and predicts characteristics of urban microclimate through the neural network enabled predictive algorithms and long-term monitoring. Two algorithms, ANN and LSTM, were selected for the predictions of characteristics of urban microclimate. Also, two prediction models were proposed to use real-time historical microclimate data with predictive techniques to predict the future real-time microclimate and abstract characteristics of microclimate. The second is to abstract characteristics of microclimate first to predict future characteristics of microclimate. Those characteristics of microclimate include the temperature and humidity (maximum, average, and minimum), cooling and heating degree days, and total amount of solar radiation in a day. About datasets, the two algorithms were fed with a long short-term urban microclimate dataset from year from year 2017 to year 2019 with 5-min resolution, and dataset from January to April 2020 was used for prediction test. From the results, in Model 1, as the time-series prediction, the ANN and LSTM algorithms can both achieve great accurate predictions of all characteristics of temperature and relative humidity, and accuracies can usually reach the best 99.92% of correlation coefficient and 98% of MAPE, for temperature, while 99.66% and 98.18% for relative humidity, respectively. In Model 2, the accuracies decreased a lot, the best accuracy of MAPE for temperature prediction can only achieve 79% while 88.6% for relative humidity, respectively. According to the prediction of solar radiation, the results show the lower accuracy, which of BP-ANN and LSTM are 51.1% and 57.8%, respectively. The results show that the more the data fluctuates, the worse the predicted results.

Author Contributions: Conceptualization, X.W. and W.W., Data curation, J.H. (Jiani Hou) and W.W., Writing—original draft, X.W., Formal analysis, Software, Validation, J.H. (Jiani Hou) and Z.T., Funding acquisition, J.H. (Jun Hui) and Z.T., Writing—review & editing, Z.T. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd. (J2021135). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of those organizations.

Data Availability Statement: We are willing to share the original data through the email of the corresponding author.

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

Appendix A

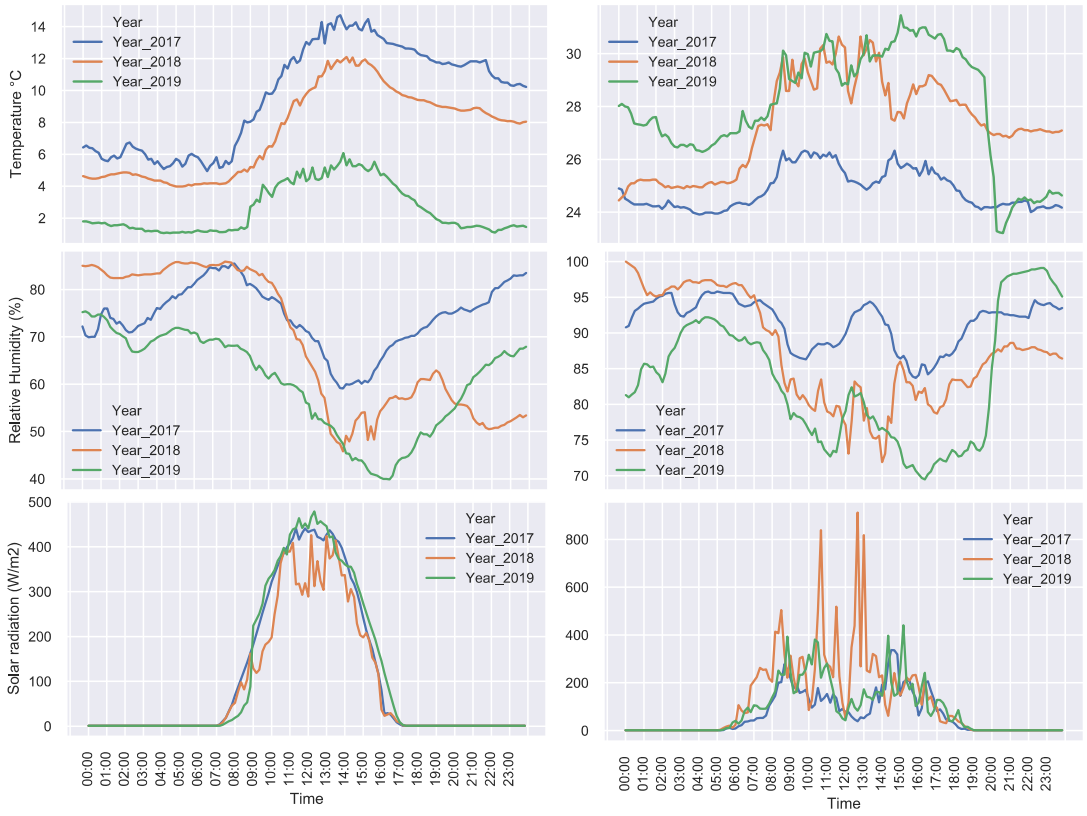


Figure A1. The example of distribution of temperature, relative humidity, and solar radiation respectively for typical days (1 January and 1 July) for three years.

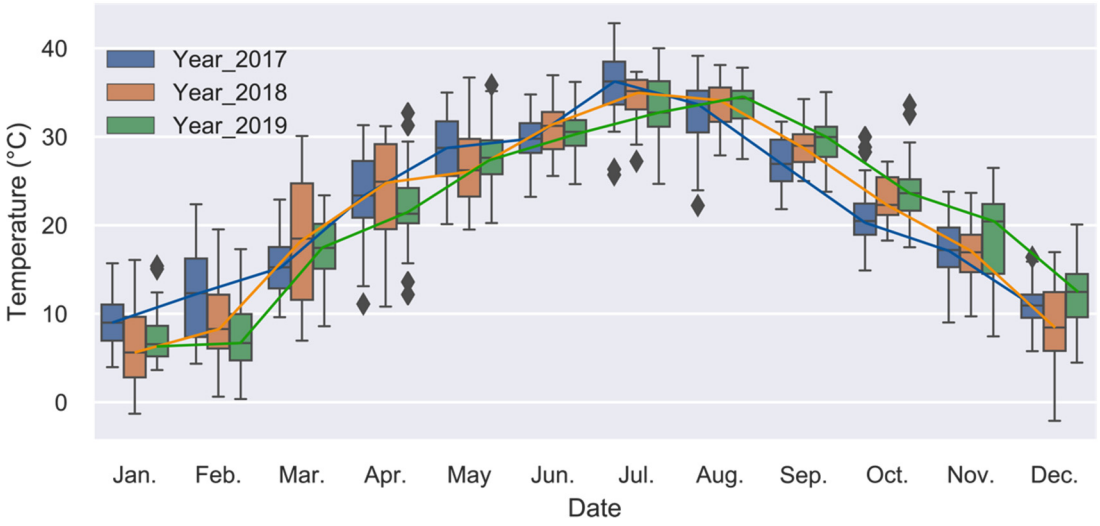


Figure A2. The distribution of maximum daily temperature for three years.

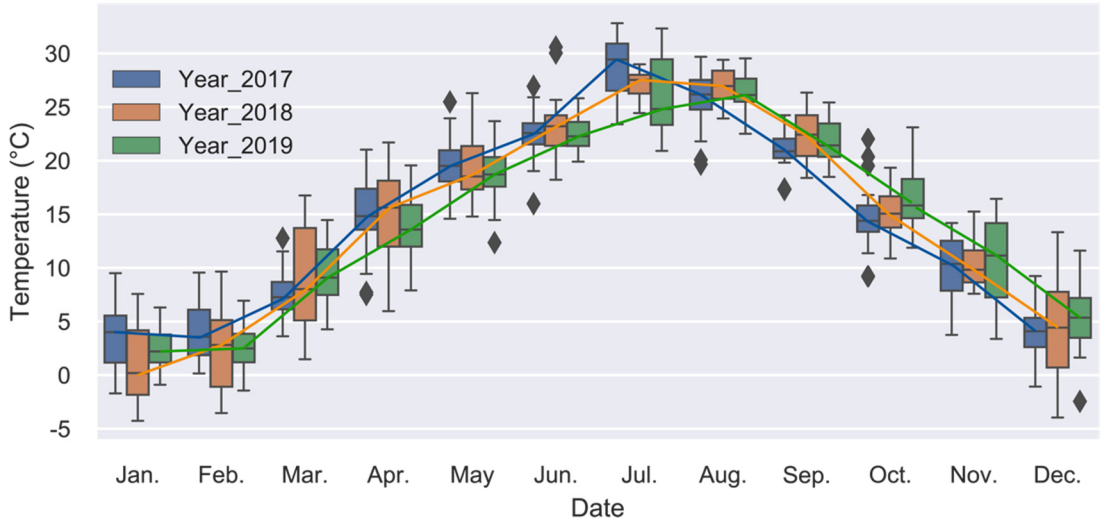


Figure A3. The distribution of minimum daily temperature for three years.

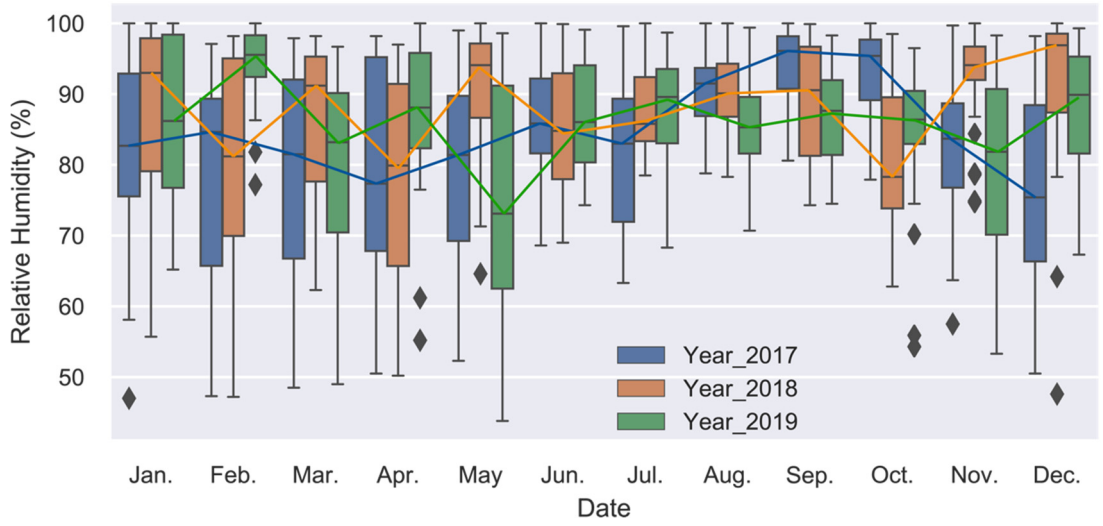


Figure A4. The distribution of maximum daily relative humidity for three years.

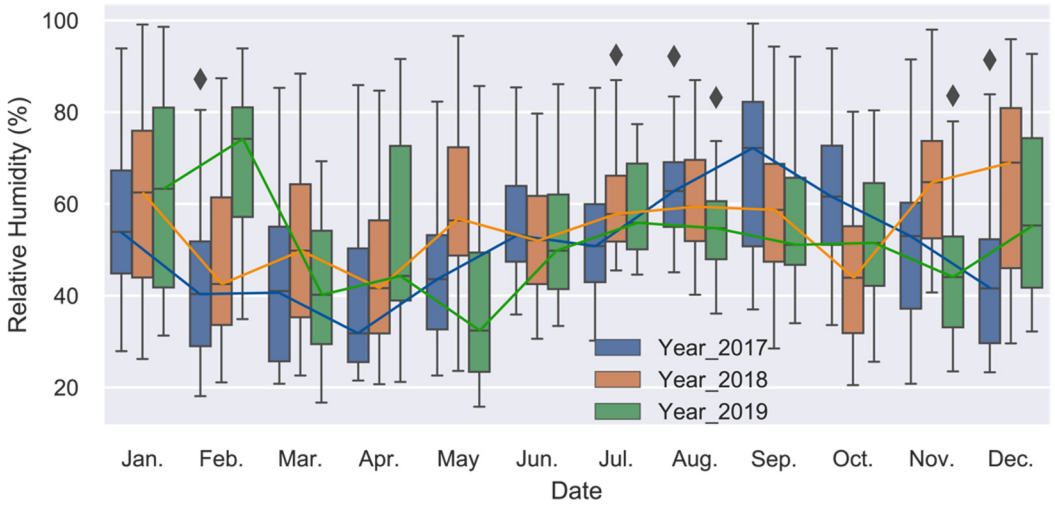


Figure A5. The distribution of minimum daily relative humidity for three years.

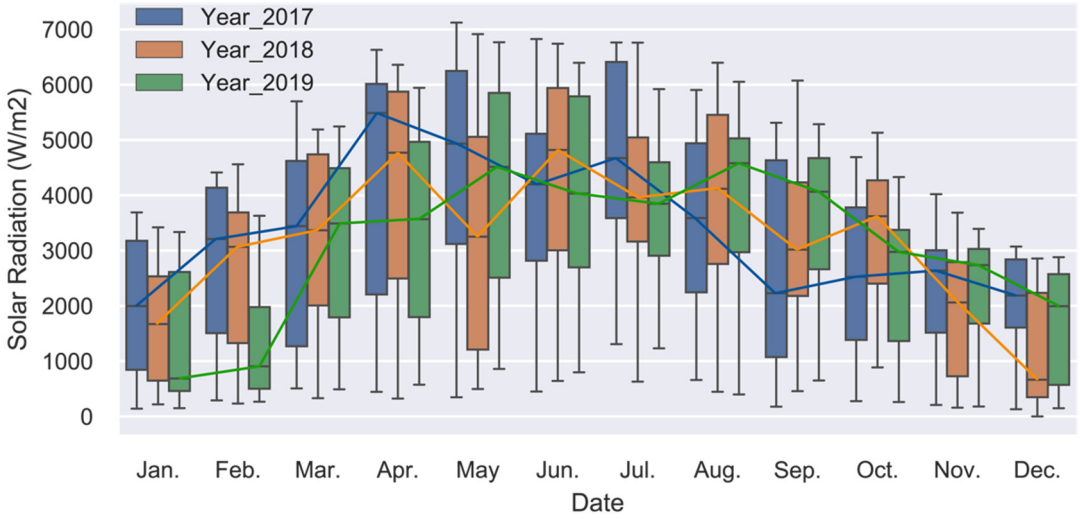


Figure A6. The distribution of sum of daily solar radiation for three years.

References

1. National Bureau of Statistics of China. *China Statistical YearBook 2021*; China Statistics Press: Beijing, China, 2021.
2. Xu, X.; Wu, Y.; Wang, W.; Hong, T.; Xu, N. Performance-driven optimization of urban open space configuration in the cold-winter and hot-summer region of China. *Build. Simul.* **2019**, *12*, 411–424. [CrossRef]
3. Elnabawi, M.H.; Hamza, N. Outdoor thermal comfort: Coupling microclimatic parameters with subjective thermal assessment to design urban performative spaces. *Buildings* **2020**, *10*, 238. [CrossRef]
4. Javanroodi, K.; Nik, V.M. Impacts of microclimate conditions on the energy performance of buildings in urban areas. *Buildings* **2019**, *9*, 189. [CrossRef]
5. Leroyer, S.; Bélair, S.; Spacek, L.; Gultepe, I. Modelling of radiation-based thermal stress indicators for urban numerical weather prediction. *Urban Clim.* **2018**, *25*, 64–81. [CrossRef]
6. Othman, N.E.; Zaki, S.A.; Rijal, H.B.; Ahmad, N.H.; Razak, A.A. Field study of pedestrians' comfort temperatures under outdoor and semi-outdoor conditions in Malaysian university campuses. *Int. J. Biometeorol.* **2021**, *65*, 453–477. [CrossRef]
7. Hayles, C.; Huddleston, M.; Chinowsky, P.; Helman, J. Quantifying the Effects of Projected Climate Change on the Durability and Service Life of Housing in Wales, UK. *Buildings* **2022**, *12*, 184. [CrossRef]
8. Zhao, Y.; Zang, Z.; Zhang, W.; Wei, S.; Xuan, Y. Predicting indoor temperature distribution based on contribution ratio of indoor climate (Cri) and mobile sensors. *Buildings* **2021**, *11*, 458. [CrossRef]
9. Heidari, H.; Mohammadbeigi, A.; Khazaei, S.; Soltanzadeh, A.; Asgarian, A.; Saghafipour, A. The effects of climatic and environmental factors on heat-related illnesses: A systematic review from 2000 to 2020. *Urban Clim.* **2020**, *34*, 100720. [CrossRef]
10. Sankar Cheela, V.R.; John, M.; Biswas, W.; Sarker, P. Combating urban heat island effect—a review of reflective pavements and tree shading strategies. *Buildings* **2021**, *11*, 93. [CrossRef]
11. Alnusairat, S.; Ayyad, Y.; Al-Shatnawi, Z. Towards meaningful university space: Perceptions of the quality of open spaces for students. *Buildings* **2021**, *11*, 556. [CrossRef]
12. Lai, D.; Chen, B.; Liu, K. Quantification of the influence of thermal comfort and life patterns on outdoor space activities. *Build. Simul.* **2020**, *13*, 113–125. [CrossRef]
13. Zhang, C.; Liao, H.; Strobl, E.; Li, H.; Li, R.; Jensen, S.S.; Zhang, Y. The role of weather conditions in COVID-19 transmission: A study of a global panel of 1236 regions. *J. Clean. Prod.* **2021**, *292*, 125987. [CrossRef] [PubMed]
14. Wang, W.; Li, S.; Guo, S.; Ma, M.; Feng, S.; Bao, L. Benchmarking urban local weather with long-term monitoring compared with weather datasets from climate station and EnergyPlus weather (EPW) data. *Energy Rep.* **2021**, *7*, 6501–6514. [CrossRef]
15. Merlier, L.; Frayssinet, L.; Johannes, K.; Kuznik, F. On the impact of local microclimate on building performance simulation. Part II: Effect of external conditions on the dynamic thermal behavior of buildings. *Build. Simul.* **2019**, *12*, 747–757. [CrossRef]
16. Buccolieri, R.; Santiago, J.L.; Martilli, A. CFD modelling: The most useful tool for developing mesoscale urban canopy parameterizations. *Build. Simul.* **2021**, *14*, 407–419. [CrossRef]
17. Bijarniya, J.P.; Sarkar, J.; Maiti, P. Environmental effect on the performance of passive daytime photonic radiative cooling and building energy-saving potential. *J. Clean. Prod.* **2020**, *274*. [CrossRef]
18. Im, H.; Srinivasan, R.S.; Maxwell, D.; Steiner, R.L.; Karmakar, S. The Impact of Climate Change on a University Campus' Energy Use: Use of Machine Learning and Building Characteristics. *Buildings* **2022**, *12*, 108. [CrossRef]
19. Bevilacqua, P.; Morabito, A.; Bruno, R.; Ferraro, V.; Arcuri, N. Seasonal performances of photovoltaic cooling systems in different weather conditions. *J. Clean. Prod.* **2020**, *272*, 122459. [CrossRef]

20. Jentsch, M.F.; James, P.A.B.; Bourikas, L.; Bahaj, A.B.S. Transforming existing weather data for worldwide locations to enable energy and building performance simulation under future climates. *Renew. Energy* **2013**, *55*, 514–524. [CrossRef]
21. Vallati, A.; Grignaffini, S.; Romagna, M.; Mauri, L.; Colucci, C. Influence of Street Canyon's Microclimate on the Energy Demand for Space Cooling and Heating of Buildings. *Energy Procedia* **2016**, *101*, 941–947. [CrossRef]
22. Wu, X.; Liu, J.; Li, C.; Yin, J. Impact of climate change on dysentery: Scientific evidences, uncertainty, modeling and projections. *Sci. Total Environ.* **2020**, *714*, 136702. [CrossRef] [PubMed]
23. Quemada-Villagómez, L.I.; Miranda-López, R.; Calderón-Ramírez, M.; Navarrete-Bolaños, J.L.; Martínez-González, G.M.; Jiménez-Islas, H. A simple and accurate mathematical model for estimating maximum and minimum daily environmental temperatures in a year. *Build. Environ.* **2021**, *197*, 107822. [CrossRef]
24. Zhang, M.; Zhang, X.; Guo, S.; Xu, X.; Chen, J.; Wang, W. Urban micro-climate prediction through long short-term memory network with long-term monitoring for on-site building energy estimation. *Sustain. Cities Soc.* **2021**, *74*, 103227. [CrossRef]
25. Moghanlo, S.; Alavinejad, M.; Oskoei, V.; Najafi Saleh, H.; Mohammadi, A.A.; Mohammadi, H.; DerakhshanNejad, Z. Using artificial neural networks to model the impacts of climate change on dust phenomenon in the Zanjan region, north-west Iran. *Urban Clim.* **2021**, *35*, 100750. [CrossRef]
26. Xie, Y.; Ishida, Y.; Hu, J.; Mochida, A. A backpropagation neural network improved by a genetic algorithm for predicting the mean radiant temperature around buildings within the long-term period of the near future. *Build. Simul.* **2022**, *15*, 473–492. [CrossRef]
27. Beroho, M.; Briak, H.; El Halimi, R.; Ouallali, A.; Boulahfa, I.; Mrabet, R.; Kebede, F.; Aboumaria, K. Analysis and prediction of climate forecasts in Northern Morocco: Application of multilevel linear mixed effects models using R software. *Heliyon* **2020**, *6*, e05094. [CrossRef]
28. Chang, J.M.H.; Lam, Y.F.; Lau, S.P.W.; Wong, W.K. Development of fine-scale spatiotemporal temperature forecast model with urban climatology and geomorphometry in Hong Kong. *Urban Clim.* **2021**, *37*, 100816. [CrossRef]
29. Papantoniou, S.; Kolokotsa, D.D. Prediction of outdoor air temperature using neural networks: Application in 4 European cities. *Energy Build.* **2016**, *114*, 72–79. [CrossRef]
30. Bueno, B.; Roth, M.; Norford, L.; Li, R. Computationally efficient prediction of canopy level urban air temperature at the neighbourhood scale. *Urban Clim.* **2014**, *9*, 35–53. [CrossRef]
31. Koc, M.; Acar, A. Investigation of urban climates and built environment relations by using machine learning. *Urban Clim.* **2021**, *37*, 100820. [CrossRef]
32. Winkler, J.; Munk, J.; Woods, J. Sensitivity of occupant comfort models to humidity and their effect on cooling energy use. *Build. Environ.* **2019**, *162*, 106240. [CrossRef]
33. Yang, Z.-c. Hourly ambient air humidity fluctuation evaluation and forecasting based on the least-squares Fourier-model. *Meas. J. Int. Meas. Confed.* **2019**, *133*, 112–123. [CrossRef]
34. Bellido-Jiménez, J.A.; Estévez Gualda, J.; García-Marín, A.P. Assessing new intra-daily temperature-based machine learning models to outperform solar radiation predictions in different conditions. *Appl. Energy* **2021**, *298*, 117211. [CrossRef]
35. Guijo-Rubio, D.; Durán-Rosal, A.M.; Gutiérrez, P.A.; Gómez-Orellana, A.M.; Casanova-Mateo, C.; Sanz-Justo, J.; Salcedo-Sanz, S.; Hervás-Martínez, C. Evolutionary artificial neural networks for accurate solar radiation prediction. *Energy* **2020**, *210*, 118374. [CrossRef]
36. Shboul, B.; AL-Arfi, I.; Michailos, S.; Ingham, D.; Ma, L.; Hughes, K.J.; Pourkashanian, M. A new ANN model for hourly solar radiation and wind speed prediction: A case study over the north & south of the Arabian Peninsula. *Sustain. Energy Technol. Assess.* **2021**, *46*, 101248. [CrossRef]
37. Pang, Z.; Niu, F.; O'Neill, Z. Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renew. Energy* **2020**, *156*, 279–289. [CrossRef]
38. Danny Harvey, L.D. Using modified multiple Heating-degree-day (HDD) and Cooling-degree-day (CDD) indices to estimate building heating and cooling loads. *Energy Build.* **2020**, *229*, 110475. [CrossRef]