Development and Comparison of Deep Learning methods for the Heating and Cooling loads Prediction of Buildings

Xiangyan Deng[1], Xinzhao Dong[1], Xu Ding[1], Zheng Yao[1], Da Li[2], Tonghui Li[3*]

[1]Zhejiang Datang Energy Marketing Co., Ltd, Hangzhou, Zhejiang Province, 310016, China
[2]China Datang Corporation Science and Technology General Research Institute Co., Ltd., East China Electric Power Test & Research Institute, Hefei, Anhui Province, 231283, China
[3]China Datang Technology Innovation Co., Ltd. Xiong'an New Area, Hebei Province, 070001, China

*Tonghui Li's Email: 278574992@qq.com

Abstract. The building industry's high energy consumption and greenhouse gas emissions make accurate heating and cooling load prediction crucial. Traditional white box and black box methods have limitations. This paper explores Deep Neural Networks (DNNs) and the Deep - FM model, which combines Factorization Machines (FM) and DNNs. The DNN imitates the human brain, while Deep - FM explicitly captures high - order feature interactions and learns non - linear patterns. Using a dataset from the UCI machine learning repository with 768 data points and 10 variables related to building attributes, the performance of these two models was evaluated. Four metrics (MAE, MSE, MAPE, and R2) were used. The model structures were optimized, and the results show that Deep - FM is more effective for simpler models and limited data, while DNN excels in complex scenarios with sufficient data. The performance of both models improves with more training data. Deep - FM outperforms DNN in predicting heating and cooling loads under small training datasets. Additionally, deep learning algorithms show much higher accuracy than shallow learning algorithms in predicting cooling load.

Keywords: factorization-machine; deep neural network; load prediction; cooling load; building attributes.

1 Introduction

The building industry significantly contributes to global energy consumption and greenhouse gas emissions, with HVAC systems accounting for about 38% of total building energy use [1]-[3]. Accurate prediction of heating and cooling loads is crucial for optimizing building structures, reducing energy consumption, and lowering investment costs, which are often inflated by calculations based on extreme conditions.

Building load prediction methods are categorized into physical (white box) models and statistical (black box) methods. White box models, such as EnergyPlus [4], DeST

© The Author(s) 2025
W. Nie et al. (eds.), *Proceedings of the 2024 6th International Conference on Hydraulic, Civil and Construction Engineering (HCCE 2024)*, Atlantis Highlights in Engineering 37,
https://doi.org/10.2991/978-94-6463-726-7_46

[5], and TRNSYS [6], require detailed building information, including structural attributes, envelope materials, equipment, and occupancy schedules. While these models are widely used for load calculation and analysis, they depend heavily on extensive input data that is often difficult to obtain accurately. Furthermore, they rely on several assumptions that may limit their applicability and lead to suboptimal prediction performance. On the other hand, black box methods, which leverage statistical and machine learning algorithms, require less detailed input data and establish predictive models through data-driven approaches. However, traditional machine learning models like decision tree (DT) [7], artificial neural networks (ANN) [8], support vector regression (SVR) [9], extreme gradient boosting (XGBoost) [10], and random forest (RF) [11] can struggle with capturing the complex, non-linear interactions between variables in building load prediction. For instance, while Seyedzadeh et al. [12] demonstrated strong RF performance using EnergyPlus datasets, and XGBoost [13]-[14] showed success, the performance of machine learning models often depends on dataset quality, input feature selection, and the algorithm used [15]-[17].

Deep neural networks (DNNs) have gained attention for their ability to learn high-order feature representations and capture complex, non-linear patterns. They address key challenges of traditional methods, such as gradient vanishing [14], and perform well in building energy prediction tasks. Models like long short-term memory (LSTM) [21], for example, have demonstrated superior performance compared to traditional shallow learning methods [22]-[25]. However, most studies utilizing DNNs focus on operational data, such as weather conditions, historical loads, and occupancy [18]-[20], with limited research on early-stage building attributes that have a strong correlation with loads during the planning phase. Although regression models have been used to estimate heating and cooling loads based on early-stage parameters [26]-[28], and studies like Tsanas [29] and Chou et al. [30] used RF and ensemble ANN+SVR models, respectively, these methods often fail to fully exploit the rich, high-dimensional relationships in building data.

To address these limitations, this paper employs the Deep-FM model, which uniquely integrates the strengths of Factorization Machines (FM) and deep neural networks. FM captures high-order feature interactions explicitly, making it particularly suitable for sparse and high-dimensional datasets often encountered in building load prediction. Meanwhile, the deep neural network component of Deep-FM learns complex, non-linear patterns directly from the data, overcoming the limitations of extensive manual feature engineering required in traditional machine learning models. Compared to white box methods, Deep-FM does not require detailed building-specific input data or rely on restrictive assumptions. Compared to black box methods, it offers improved scalability, robustness, and predictive accuracy by leveraging both explicit feature interaction modeling and implicit representation learning.

In this study, we optimize and compare the performance of DNN and Deep-FM models for predicting building loads. By investigating the differences between shallow machine learning and advanced deep learning algorithms, we demonstrate the superiority of Deep-FM in handling high-dimensional, sparse, and non-linear data, making it better suited for building load prediction tasks than traditional methods.

2 Methods and Materials

2.1 DNN and Deep-FM

The behavior of DNN imitates the human brain and receives raw data to make intelligent decisions. The neural network is consist of many interconnected neurons, which is used to deal with input parameters and obtain the optimum target. An entire DNN commonly contains input layer, hidden layers and output layer, which the number of hidden layers is more than 2. The relationship between input layer and output layer in DNN architecture is shown as:

$$y_{DNN} = G(\gamma_i + A_i X_i) \tag{1}$$

$$G(x) = \max(0, x) \tag{2}$$

Where y_{DNN} is the output of DNN model. i is the ith layer of DNN. Ai is the weight matrix of ith layer, γi is the bias matrix of the ith layer. G is the activation function of hidden layer, which includes sigmoid, tanh, rectified linear unit (ReLU) etc (Eq.2). The ReLU is a non-linear transfer function and adopted as activation function of DNN hidden layers in our paper due to fast convergence and overcoming gradient vanishing.

Factorization-machine based neural network (Deep-FM) is new deep learning code developed by Huawei Company, which integrates deep neural network (DNN) and factorization machine (FM) [26]. The aim of FM is to learn both low and high order feature interactions. Taking FM component and DNN component into account, the combined model is trained jointly as:

$$y_{Deep-FM} = Sigmoid(y_{FM} + y_{DNN}) \tag{3}$$

Where $y_{Deep\text{-}FM}$, y_{FM}, y_{DNN} is the output of Deep-FM, FM component, DNN component respectively. The output of FM is calculated as:

$$y_{FM} = \gamma_{FM} + A_{FM} x_0 + \sum_j \sum_i x_{0,i} x_{0,j} (U_i U_j) \tag{4}$$

Where γ_{FM} is the bias matrix of FM component, A_{FM} is the weight matrix of FM component. x is the input parameters. U is the hidden matrix reflecting input variable interactions. The deep neural network has an enormous potential on small datasets compared with a huge dataset with millions or even billions data [31]. The adequate learning rate for DNN and Deep-FM is 0.001. The adopted epoch for DNN and Deep-FM is 100.The optimization of the structure for both DNN and Deep-FM is detailed in section 4.1. The DNN model and Deep-FM model are implemented based upon DeepCTR with TensorFlow backend.

Four metrics are used to evaluate the performance of models, mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), and coefficient of determination (R2). To evaluate the performance of model accurately, we conducted ten times experiments independently and adopted the average values of four metrics.

2.2 Dataset and Data Preprocessing

The Dataset in our paper derived from UCI machine learning repository. The detailed description of dataset is shown in the literature [29]. Briefly, the physical models of building were generated by Ecotect software. The buildings locate in Athens, Greece. The buildings with the same volume 771.75 m^2. The building used the most common materials with the lowest U-value, as follows: walls (1.78), floors (0.86), roofs (0.5), and windows (2.26) [29]. The glazing area is expressed as the percentages of floor area. The detailed explanation about orientation and glazing area distribution is shown in literature [29].

Table 1. Dataset description

	Maximum	Minimum	Average	Median	Standard deviation
Relative compactness	0.98	0.62	0.7642	0.75	0.1058
Surface area	808.5	514.5	671.708	673.75	88.0861
Wall area	416.5	245.0	318.5	318.5	43.6265
Roof area	220.5	110.25	176.6042	183.75	45.166
Overall height	7.0	3.5	5.25	5.25	1.7511
Orientation	5.0	2	3.5	3.5	1.119
Glazing area	0.4	0	0.2344	0.25	0.1332
Glazing area distribution	5	0	2.81	3.0	1.551
Heating load	43.1	6.01	22.3072	18.95	10.09
Cooling load	48.03	10.9	24.5878	22.08	9.5133

The dataset contains 768 data points and 10 variables. Table 1 shows the statistical analysis of the building attributes and heating/cooling loads, which contains max value, min value, median value and standard deviation of data. The input variables are consist of relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area and glazing area distribution. The output variables are heating and cooling load. The input parameters belong to sparse feature. For instance, glazing area in dataset has three types: 0, 0.1, 0.25 and 0.4 of floor area.

3 Result

3.1 Model Structure Optimization

Structure optimizations for both DNN and Deep-FM were conducted separately. The evaluation indicators of various structures for both DNN and Deep-FM on predicting heating and cooling load are presented in Tables 2 and 3. Mean absolute error, mean squared error, mean absolute percentage error, coefficient of determination are used as evaluating metrics. For heating load, Deep-FM model with three hidden layers and 128 neurons in each hidden layer (in bold in Table 2) shows a better prediction performance since all metrics remain to a better level. And DNN model with two hidden layers and 128 neurons per layer (in bold in Table 2) has a better prediction performance. In addition, the optimized Deep-FM model presents lower metrics and better predicting performance compared with DNN model.

Table 2. Metrics of heating load for various structures of the model

Hidden layer	**Neuron number/layer**	**Deep-FM**				**DNN**			
		MSE	MAE	MAPE	R^2	MSE	MAE	MAPE	R^2
2	16	0.3374	0.3958	1.78	0.9967	2.5263	0.8090	3.79	0.9760
	32	0.2970	0.3821	1.71	0.9970	1.3539	0.5363	2.52	0.9868
	64	0.2661	0.3599	1.58	0.9975	0.2767	0.3905	1.85	0.9974
	128	0.2380	0.3327	1.50	0.9976	**0.2880**	**0.3780**	**1.78**	**0.9971**
3	16	0.3483	0.3935	1.74	0.9966	0.8424	0.4899	2.20	0.9918
	32	0.2552	0.3505	1.51	0.9974	0.4216	0.4506	2.09	0.9958
	64	0.2862	0.3664	1.64	0.9972	0.3152	0.4037	1.91	0.9969
	128	**0.2319**	**0.3304**	**1.40**	**0.9978**	0.4058	0.4277	1.86	0.9961
4	16	0.2730	0.3641	1.58	0.9973	0.7930	0.4937	2.23	0.9922
	32	0.2713	0.3551	1.56	0.9971	0.3097	0.3951	1.86	0.9968
	64	0.2927	0.3719	1.66	0.9972	0.3205	0.3965	1.87	0.9969
	128	0.2718	0.3546	1.52	0.9975	0.3048	0.3922	1.77	0.9972
5	16	0.3232	0.3921	1.77	0.9970	0.5200	0.5014	2.62	0.9952
	32	0.2485	0.3454	1.55	0.9976	0.3258	0.4289	2.09	0.9969
	64	0.2502	0.3522	1.53	0.9975	0.3385	0.4176	1.90	0.9967
	128	0.3444	0.3848	1.72	0.9965	0.3069	0.3995	1.90	0.9969

Table 3. Metrics of cooling load for various structures of the model

Hidden layer	**Neuron number/layer**	**Deep-FM**				**DNN**			
		MSE	MAE	MAPE	R^2	MSE	MAE	MAPE	R^2
2	16	4.7377	1.0765	3.67	0.9474	6.7728	1.4888	5.27	0.9250
	32	4.7963	1.0538	0.0361	0.9486	3.2384	0.9480	0.0332	0.9653
	64	3.5843	1.0336	0.0352	0.9591	3.0435	1.0529	0.0367	0.9651
	128	4.2998	1.0853	0.0373	0.9500	1.4864	0.8444	0.0326	0.9827
3	16	3.2099	0.9001	0.0317	0.9636	4.8070	1.2764	0.0465	0.9471
	32	2.2281	0.8441	0.0298	0.9734	2.5734	0.9823	0.0345	0.9692
	64	2.5016	0.8081	0.0285	0.9719	2.6505	0.9200	0.0325	0.9702
	128	0.6682	0.6072	0.0258	0.9925	0.8448	0.6286	0.0246	0.9902
4	16	5.2937	1.1268	0.0387	0.9429	2.5384	0.9864	0.0349	0.9727
	32	2.8436	0.9658	0.0338	0.9662	4.7927	1.5433	0.0572	0.9430
	64	1.6833	0.7765	0.0297	0.9813	1.6833	0.7765	0.0297	0.9813
	128	**0.5883**	**0.5383**	**0.0232**	**0.9936**	0.6546	0.5258	0.0213	0.9929
5	16	4.5000	0.9815	0.0342	0.9485	3.6520	1.1455	0.0407	0.9585
	32	2.4893	0.9756	0.0335	0.9742	2.7328	1.0401	0.0355	0.9717
	64	1.8413	0.7768	0.0276	0.9776	0.5719	0.5293	0.0221	0.9930
	128	0.6528	0.5242	0.0207	0.9925	**0.4324**	**0.4721**	**0.0186**	**0.9951**

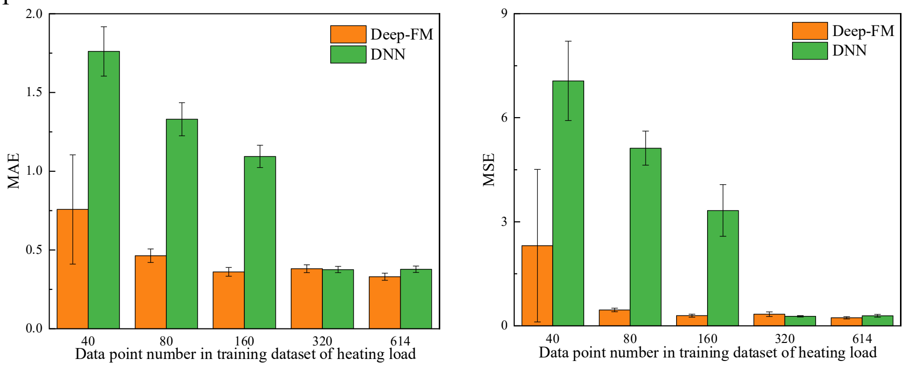
The performance of Deep-FM and DNN models was compared using MSE, MAE, MAPE, and R^2 across different configurations of hidden layers and neurons per layer. Deep-FM outperformed DNN in simpler configurations, such as two-layer models with 16 neurons per layer, achieving an MSE of 4.7377 and MAE of 1.0765, compared to DNN's MSE of 6.7728 and MAE of 1.4888. However, at higher neuron counts (e.g., 128 neurons), DNN performed better with an MSE of 1.4864 and R2R^2R2 of 0.9827, highlighting its advantage with larger architectures.

For three-layer models, Deep-FM excelled at 128 neurons, achieving its best performance with an MSE of 0.6682 and R^2 of 0.9925, outperforming DNN's MSE of 0.8448 and R^2 of 0.9902. In four- and five-layer configurations, DNN surpassed Deep-FM as neuron counts increased, achieving its best results at five layers with 128 neurons, where it had an MSE of 0.4324 and R2R^2R2 of 0.9951, compared to Deep-FM's MSE of 0.6528 and R^2 of 0.9925.

Overall, Deep-FM is more effective for simpler models and limited data, while DNN excels in complex scenarios with sufficient data and computational power. The choice of model depends on the specific requirements of the prediction task.

4 Discussion

The comparisons of the performance of both DNN and Deep-FM model was conducted under different data point. Fig.1 shows the MSE and MAE of both DNN and Deep-FM models in predicting heating load. Notably, the advantage of Deep-FM model in predicting heating load is very obvious compared with DNN model when the training data point below 160. When the training data number above or equal 320, the MAE and MSE of DNN model are close to those of Deep-FM model. Additionally, the error of MAE and MSE decrease with the increase of data point. However, The error of MSE and MAE of Deep-FM model is far more than other error when training data point is 40, even reach to 100%. It indicated that the larger data point there is, the better stable prediction will be.



(a) (b)

Fig. 1. Comparison of the metrics of DNN and Deep-FM for predicting heating load. (a) MAE; (b) MSE (the results were calculated based on the test dataset).

The MAE and MSE of models for predicting cooling load are presented in Fig.2. For DNN model, the MAE and MSE reduce as the increase of training data point. It indicated that the accuracy of DNN model depends on the training data size. For Deep-FM model, when training data point below 320, the data number has low impact on the accuracy. The values of MAE and MSE have an outstanding reduction when training data number is 614. Moreover, the error of MSE of DNN model is also larger than those of Deep-FM model in predicting cooling load. The error of MSE and MAE of Deep-FM model in predicting cooling load is far less than that in predicting heating load when training data point is 40. When training data point is 40 or 80, Deep-FM model has better performance. As the amount of training data point increases, the accuracy of DNN model exceed Deep-FM model.

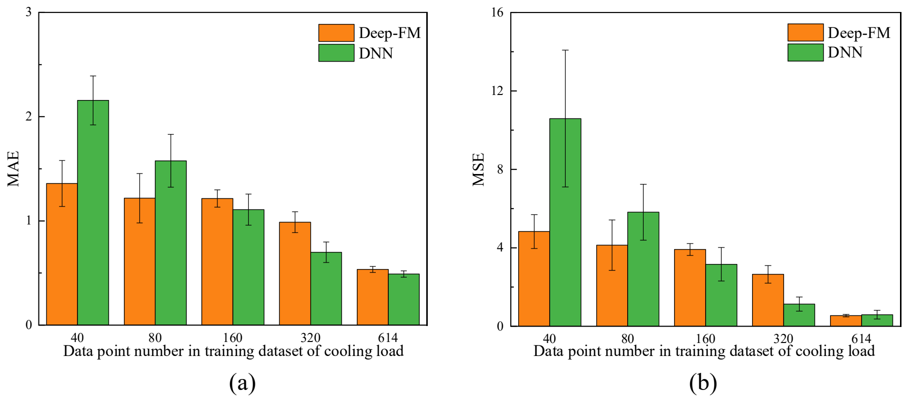


Fig. 2. Comparison of the metrics of DNN and Deep-FM for predicting cooling load. (a) MAE; (b) MSE (the results were calculated based on the test dataset).

5 Conclusions

In conclusion, this study applied DNN and Deep - FM, two deep - learning algorithms, to predict heating and cooling loads using a building attributes dataset. Multiple building parameters, including relative compactness, surface area, wall area, and roof area, were utilized as input variables. The architectures of the two deep - learning models were optimized, and the influence of dataset size on prediction performance was explored. The findings demonstrate a significant improvement in model prediction performance with the growth of the number of training data. Additionally, a comparison of the performance of Deep - FM and DNN models under different numbers of training data points was conducted. The results suggest that the Deep - FM model performs better in predicting heating and cooling loads when the amount of training data is small. Finally, a contrast of the numerical evaluations between shallow - learning models (ANN, SVR, CART, CHAID, GLR, RF, IRLS, and ELM) and deep - learning models (DNN and Deep - FM) was made. Evidently, in predicting cooling load, the accuracy of deep - learning algorithms is much higher than that of shallow - learning algorithms.

This research provides valuable insights into the application of deep - learning algorithms in predicting building heating and cooling loads, and offers a reference for further optimization of energy - related predictions in the building field.

References

1. González-Torres, M.; Pérez-Lombard, L.; Coronel, J. F.; Maestre, I. R.; Yan, D. A review on buildings energy information: Trends, end-uses, fuels and drivers. Energy Rep 2022, 8, 626-637.
2. Zhao, X.; Liu, S.; Yin, Y.; Zhang, T. T.; Chen, Q. Airborne transmission of COVID-19 virus in enclosed spaces: An overview of research methods. Indoor Air 2022, 32, e13056.
3. Zhao, X.; Yin, Y. Inverse regulation of the indoor environment: An overview of the adjoint method. Energ Buildings 2022, 259, 111907.
4. Crawley, D. B.; Lawrie, L. K.; Winkelmann, F. C.; Buhl, W.F.; Huang, Y. J.; Pedersen, C. O.; Strand, R. K.; Liesen, R. J.; Fisher, D. E.; Witte, M. J.; Glazer, J. EnergyPlus: creating a new-generation building energy simulation program. Energ Buildings 2001, 33, 319-31.
5. Yan, D.; Xie, X.; Song, Y. F. Building environment design simulation software DeST (1): an overview of developments and information of building simulation and DeST. J HVAC 2004, 34, 48-56
6. Klein, S.A.; Beckman, W. A.; Duffie, J. A. Trnysys - a transient simulation program. Build Eng 1976, 82.
7. Quinlan, J.R. Induction of decision trees. Mach Learn 1986, 1, 81-106.
8. Hecht-Nielsen, R. Theory of the backpropagation neural network, Neural Networks, 1992.
9. Suykens, J. A. K.; Vandewalle, J. Least squares support vector machine classifiers. Neural Process Lett 1999, 9, 293-300.
10. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. ACM SIGKDD and International Conference on Knowledge Discovery and Data Mining, 2016.
11. Liaw, A.; Wiener, M. Classification and regression by randomForest. R News 2002, 23, 18-22.
12. Seyedzadeh, S.; Rahimian, F. P.; Oliver, S.; Glesk, I.; Kumar, B. Data driven model improved by multi-objective optimisation for prediction of building energy loads. Autom Con-Struct 2020, 116, 103188.
13. Al-Rakhami, M.; Gumaei, A.; Alsanad, A.; Alamri, A.; Hassan, M. M. An ensemble learning approach for accurate energy load prediction in residential buildings. IEEE Access 2019, 7, 48328-48338.
14. Wang, Z.; Hong, T.; Piette, M. A. Building thermal load prediction through shallow machine learning and deep learning. Appl Energy 2020, 263, 114683.
15. Zhou, Y.; Liu, Y.; Wang, D.; Liu, X. Comparison of machine-learning models for predicting short-term building heating load using operational parameters. Energ Buildings 2021, 253, 111505.
16. Ngo, N. T. Early predicting cooling loads for energy-efficient design in office buildings by machine learning, Energ Buildings 2019, 182, 264-273.
17. Luo, X. J.; Oyedele, L. O.; Ajayi, A. O.; Akinade, O. O. Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads. Sustain Cities Soc 2020, 61, 102283.
18. Lei, L.; Chen, W.; Wu B.; Chen, C.; Liu, W. A building energy consumption prediction model based on rough set theory and deep learning algorithms, Energ Buildings 2021, 240, 110886.

19. Bedi, J., Toshniwal, D. Deep learning framework to forecast electricity load. Appl Energy 2019, 238, 1312-1326.
20. He, W. Load forecasting via deep neural networks. Procedia Computer Science 2017, 122, 308-314.
21. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput 1997, 9,1735-1780.
22. Wang, Y.; Velswamy, K.; Huang, B. A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems. Processes 2017, 5, 46.
23. Zhou, C.; Fang, Z.; Xu, X.; Zhang, X.; Ding, Y.; Jiang, X. Using long short-term memory networks to predict energy consumption of air-conditioning systems. Sustain Cities Soc 2020, 55, 102000.
24. Zheng, Z.; Chen, H. N.; Luo, X. W. A Kalman filter-based bottom-up approach for household short-term load forecast. Appl Energy 2019, 250, 882–894.
25. Wei, Z.; Zhang, T.; Yue, B.; Ding, Y.; Xiao, R.; Wang, R.; Zhai. X. Prediction of residential district heating load based on machine learning: A case study. Energy 2021, 231, 120950.
26. Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. Proceedings of the 26th International Joint Conference on Artificial Intelligence 2017.
27. Korolija, I.; Zhang, Y.; Marjanovic-Halburd, L.; Hanby V. I. Regression models for predicting UK office building energy consumption from heating and cooling demands. Energ Buildings 2013, 59, 214-227.
28. Catalina, T.; Virgone, J.; Blanco, E. Development and validation of regression models to predict monthly heating demand for residential buildings. Energ Buildings 2008, 40, 1825-1832.
29. Tsanas, A.; Xifara, A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, Energ Buildings 2012, 49, 560-567.
30. Chou, J. S.; Bui, D. K. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energ Buildings 2014, 82, 437-446.
31. Zhang, X. C.; Gong, J. G.; Xuan, F. Z. A deep learning based life prediction method for components under creep, fatigue and creep-fatigue conditions. Int J Fatigue 2021,148, 106236.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.