Proceedings of the 7h World Congress on Civil, Structural, and Environmental Engineering (CSEE'22)
~~Lisbon, Portugal~~ Virtual Conference – April 10 – 12, 2022
Paper No. ICSECT 174
DOI: 10.11159/icsect22.174

A Review of Machine Learning Methods in Building Energy Performance Assessment

Handan Kunkcu[1], Kerim Koc[2], Asli Pelin Gurgun[2]
[1]Eskisehir Technical University
Department of Civil Engineering, Eskisehir, Turkey
handan424@eskisehir.edu.tr; kerimkoc@yildiz.edu.tr;
[2] Yildiz Technical University
Department of Civil Engineering, Istanbul, Turkey
apelin@yildiz.edu.tr

Abstract **-** Buildings alone account for nearly one third of the world's total energy consumption. Therefore, it is of great importance to determine the energy performance of buildings with prediction models to contribute to the reduction in energy consumption. Machine learning (ML) methods have extensively been adopted in the assessment of the energy performance of buildings in the literature. This research aims to review existing studies in the literature to predict energy performance through ML methods. According to a comprehensive literature survey, 79 articles were identified and reviewed intensively based on several aspects such as journal, publication year, examined country, adopted programming tools, building types and utilized ML methods. Results of the literature survey indicated that different building types (i.e., residential or non-residential buildings) have different energy performances. Methodological investigation showed that artificial neural network and support vector machine methods were the most frequently implemented ML techniques in the prediction of energy performance. It was observed that many authors compared the performances of several ML methods to highlight the most capable methods. In addition, the energy performances of buildings were evaluated with ML methods using different programming tools in variety of countries. Overall, this study is expected to provide valuable information about the current state of ML methods to practitioners and researchers in this field.

Keywords**:** Building, energy performance, artificial intelligence, prediction, review, energy consumption, facility management.

1. Introduction

Globally, energy plays an important part in the development of countries [1]. Buildings account for around 30% of total energy consumption worldwide. Hence, forecasting the energy consumption of buildings, implementing efficient energy management, and improving the energy performance of current buildings are vital for building energy conservation [2]. Additionally, energy performance of buildings could be significantly influenced by climatic environment and specific information of building systems (i.e., building equipment, building size, occupant behavior, indoor environment) [3], [4]. Since assessment of the energy performance of buildings relies on many parameters, machine learning (ML) methods can be used to provide prediction outcomes for the early designation of building performance by researchers and/or practitioners. ML methods have widely been adopted in the literature to present a futuristic basis for complicated prediction problems [5]. In this context, energy performance of buildings has been investigated using various machine learning methods, such as artificial neural network (ANN) [6], random forest (RF) [5], support vector machine (SVM) [7], decision tree (DT) [8], k-nearest neighbor (KNN) [9], extreme gradient boosting (XGBoost) [3], and deep neural network (DNN) [10].

The purpose of this study is to review existing studies on ML applications in the assessment of the energy performance of buildings. The articles attained via the Scopus search engine were investigated with the following features: journal name, publication year, examined country, used programming tools, building types, and adopted ML methods. The findings of this study can provide practitioners with a conceptual framework to help them perform energy analysis through ML methods, with a particular emphasis on the past investigations. It is believed that different combinations of ML methods (ensemble approaches) [11] or new ML methods [12], [13] will be developed in the future, in a way that is directed by review studies highlighting applicability of past attempts.

2. Methodology

A review on the energy performance of buildings was carried out via the Scopus search engine in the first step. A total of 146 publications was obtained from the initial search using the following search code: TITLE-ABS-KEY ("artificial intel*" OR "machine learning" OR "artificial neural network" OR "random forest" OR "support vector machine" OR "deep neural network" OR "deep learning" OR "XGBOOST" OR "Naive Bayes*" OR "decision tree" OR "Bayesian Network" OR "Adaptive Boosting" OR "Empirical Bayes" OR "gradient boosting algorithm" OR "support vector regression" OR "ANN" OR "SVM" OR "SVR" OR "SVC" OR "RF" OR "DNN" OR "ML" OR "AI" OR "BN" OR "NB" OR "KNN" OR "GBM") AND TITLE-ABS-KEY ("energy performance" OR "energy analysis") AND TITLE-ABS-KEY (building)) AND (LIMIT-TO (SRCTYPE,"j")) AND (LIMIT-TO (SUBJAREA,"ENGI")) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"re")). In addition, the search results were limited to journal articles between 2017 and 2022, which were written in English. This step was followed by the abstract level investigation, to ensure that the included studies were in line with the study objective regarding building energy performance and ML techniques. Accordingly, 79 research papers were found to be appropriate for the objective of this study.

3. Results and Discussion

Figure 1 presents the percentages of the published articles according to journals. It was found that 22% of the studies have been published in the *Energy and Buildings*, followed by *Applied Energy*, and *Energy* with 15% and 14%, respectively. Some of the journals in the "Other" that corresponds to the 34% of reviewed studies were *Automation in Construction, Building and Environment, Journal of Building Engineering, Journal of Building Performance Simulation, Journal of Cleaner Production*.

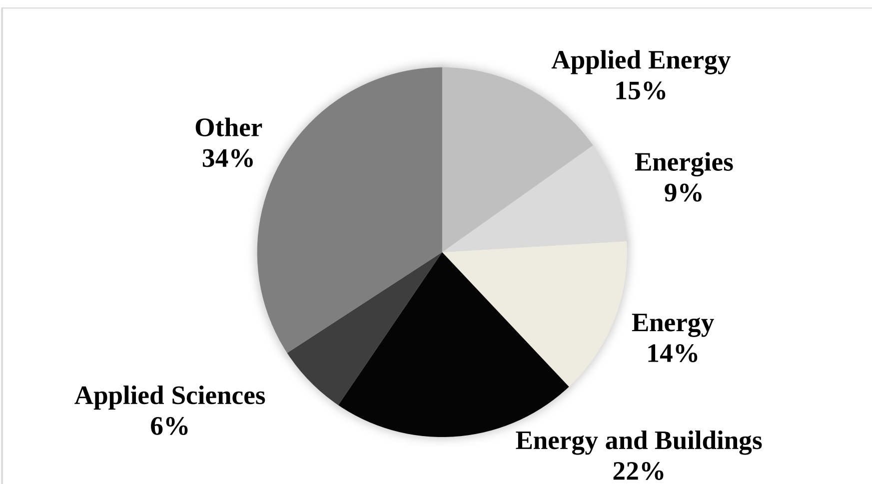

Fig. 1: Percentage distribution of selected articles according to journal.

Figure 2 shows the distribution of articles by publication year. An increasing trend can be easily followed from the figure, such that the number of published works reached its maximum in 2020 and 2021 with 20 and 23 research articles. As the building's energy performance is based on the climatic environment and specific information of the buildings, holistic and diverse approaches were adopted in the studies. In addition, building information modelling (BIM), optimization, GIS-based modelling and fuzzy approaches were also adopted along with ML methods to boost the robustness of the methodological approaches, particularly in the articles published in 2020 and 2021.

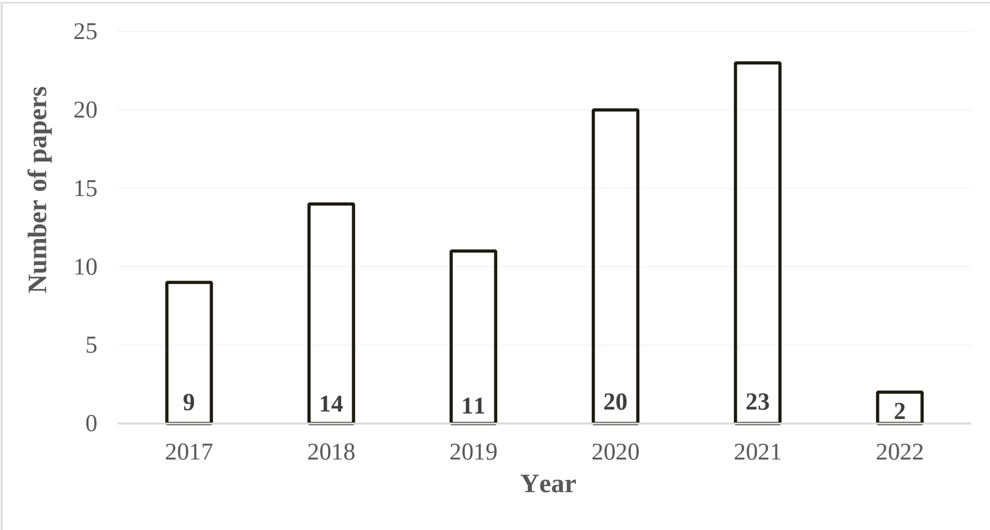

Fig. 2: Distribution of the selected articles according to year.

Table 1 provides the distribution of publications according to country. It was observed that ML applications were conducted in a variety of countries in the reviewed articles. The differences in the regulations and standards of the countries about the building energy performance were frequently discussed in the publications and performance analysis of buildings were performed in several countries. For example, dynamic simulation models were developed for Germany, Spain, the United Kingdom, Belgium, Italy, France, and Sweden [14]. When the research articles between 2017 and 2022 were examined, more case studies have been conducted in developed countries compared to developing countries, showing that attaining energy-related data and cases were easier in developed countries.

Table 1: Distribution of examined countries in the selected articles.

Country	Number of articles
US	7
UK	6
Canada	5
China	5
Germany	5
Italy	5
Spain	5
Korea	4
Greece	3
Netherlands	3
Others	19
N/A	24

It has also been observed that energy and cost performance of the buildings showed significant alterations based on the type of the buildings. Table 2 represents the distribution of building types such as residential buildings and non-residential buildings. Residential buildings were family house buildings, single-person houses and dwellings and non-residential buildings included hospitals, schools, office buildings, hotels, stores, warehouses, strip malls, restaurants, libraries, museums, congress halls, gymnasium and sports buildings, laboratories, industrial buildings in the reviewed publications. Since non-residential buildings such as laboratories and dormitories are energy-intensive areas with varying electrical load densities

[15], they were more preferred in the assessments of energy performance. In addition, the energy programming tools used to develop ML models are listed in Table 3. MATLAB was mostly integrated with EnergyPlus and TRNSYS software in many articles because it provides the management and optimization of simulations [16]. EnergyPlus was the most preferred building energy simulation software in the articles. As researchers in [16] pointed out, the reason for preference might be related to that EnergyPlus achieves high accuracy in the evaluation of dynamic energy performance.

Table 2: Distribution of the building type in the selected articles.

Type of buildings	Number of articles	Percentage
Residential buildings	29	27
Non-residential buildings	70	66
N/A	7	7

Table 3: Building energy performance forecasting using software and programming language tools.

Software and programming language tools	Number of articles
MATLAB	16
EnergyPlus	15
Python	10
R environment support	10
Ecotect	7
TRNSYS	4
IES-VE	3
Others	12

Table 4 shows the summary of the ML methods adopted in estimating the building energy performance. The ML methods adopted in the reviewed articles were categorized as tree-based, neural network-based, vector-based, distance-based and probability-based methods. It is noteworthy to mention that ANN, RF and SVM were examined in 37, 21 and 19 of the studies, respectively. The pertinent literature [17], [18] supports the results attained in this study. Accordingly, ANN and SVM methods were extensively applied in building energy forecasting models and building energy performance assessments. Compared to DT and other algorithms, ANN and SVM has been utilized more commonly due to the higher accuracy in prediction despite they require many parameters, to be tuned [19]. On the other hand, the RF method was preferred in numerous articles because it provides high prediction performance in large data sets [20], as well as presenting high interpretability of the model outputs.

Table 4: Numeric distribution of ML methods in the selected articles.

ML methods		Abbreviation	Number of articles
Tree based ML methods	Random Forest	RF	21
	Extreme Gradient Boosting	XGBoost	8
	Gradient Tree Boosting	GTB	7
	Decision Tree	DT	6
	Classification and Regression Tree	CART	2
	Alternating Model Tree	AMT	1
	Boosted Tree	BT	1
	Categorical Boosting	CatBoost	1
	Extremely Randomized Tree or Extra Tree	ET	1
Neural network based ML methods	Artificial Neural Network	ANN	37
	Deep Neural Network	DNN	10
	Multi-layer Perceptron Neural Network	MLPNN	7
	Neural Network	NNet	3
	Recurrent Neural Network	RNN	3
	Adaptive Neuro-Fuzzy Inference System	ANFIS	2
	Bayesian Neural Network	BNN	2
	Convolutional Neural Network	CNN	2
	Extreme Learning Machine	ELM	2
	Long Short Term Memory	LSTM	2
	Radial Basis Function Network	RBFN	2
	Multi-layer Feedforward Neural Networks	MFNN	1
	Nonlinear Autoregressive Exogenous Recurrent Neural Network	NARX RNN	1
	Residual Network	ResNet	1
Vector based ML methods	Support Vector Machine	SVM	19
Distance based ML methods	K-Nearest Neighbor	KNN	2
Probability based ML methods	Bayesian Network	BN	2
	Naive Bayes	NB	1
Other	Linear Regression/Logistic Regression	LR	12

ML methods can solve both classification and regression problems. Figure 3 indicates the percentages of the addressed problem type in examined articles. Accordingly, majority of the researchers addressed regression problems (88%), while classification problems were considered in only 12% of the studies.

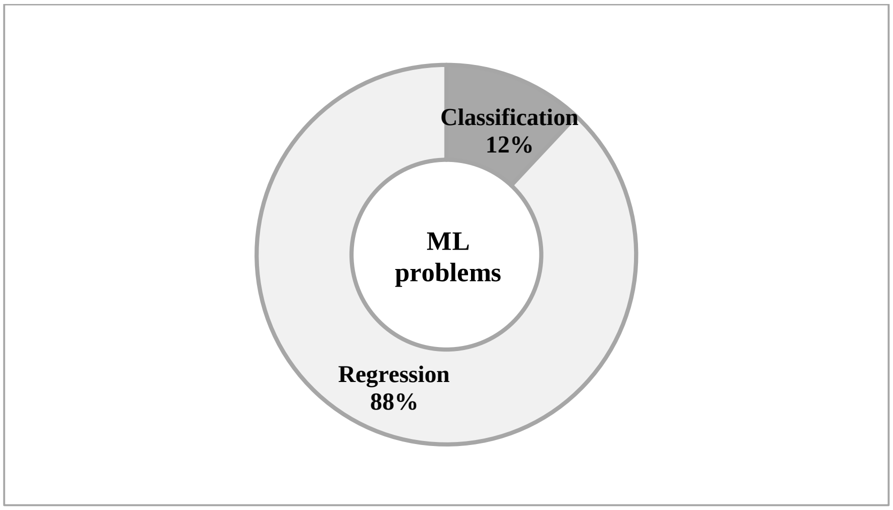

Fig. 3: Distribution of the selected articles according to developed models.

4. Conclusion

Investigating the energy performance of buildings is essential to early detection of building energy performance trends. Since the energy performance of buildings depends on many factors, the evaluation of energy performance can be considered as a complex and laborious process. In this study, it was found that ML methods provide a great convenience in the performance-based design, modelling, and analysis of buildings. This study aims to investigate the trends of ML applications in building energy performance. The findings show that predictive methods have widely been implemented in the literature to estimate the energy performance of large-scale buildings, such as public buildings and office buildings. Most of the researchers evaluated and tested the proposed methods as a case study in different countries. While ANN, RF and SVM were the most commonly used ML methods, CatBoost, the newly developed method, was rarely adopted. This study presents a detailed analysis of current ML programming tools and used methods. The ML models can be used for the early planning of productivity loss in the energy performance of buildings. Hence, this study can contribute to the development of an effective ML model regarding the building energy performance analysis field in future studies.

References

[1] H. Tabrizchi, M. M. Javidi, and V. Amirzadeh, "Estimates of residential building energy consumption using a multi-verse optimizer-based support vector machine with k-fold cross-validation," *Evol. Syst.*, vol. 12, no. 3, pp. 755–767, 2021, doi: 10.1007/s12530-019-09283-8.

[2] S. A. Sharif and A. Hammad, "Developing surrogate ANN for selecting near-optimal building energy renovation methods considering energy consumption, LCC and LCA," *J. Build. Eng.*, vol. 25, no. April, p. 100790, 2019, doi: 10.1016/j.jobe.2019.100790.

[3] Y. Feng, Q. Duan, X. Chen, S. S. Yakkali, and J. Wang, "Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods," *Appl. Energy*, vol. 291, no. December 2020, p. 116814, 2021, doi: 10.1016/j.apenergy.2021.116814.

[4] Y. R. Yoon and H. J. Moon, "Energy consumption model with energy use factors of tenants in commercial buildings using Gaussian process regression," *Energy Build.*, vol. 168, pp. 215–224, 2018, doi: 10.1016/j.enbuild.2018.03.042.

[5] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, and J. Wang, "A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning," *Appl. Energy*, vol. 235, no. November 2018, pp. 1551–1560, 2019, doi: 10.1016/j.apenergy.2018.11.081.

[6] B. Chegari, M. Tabaa, E. Simeu, F. Moutaouakkil, and H. Medromi, "Multi-objective optimization of building energy

performance and indoor thermal comfort by combining artificial neural networks and metaheuristic algorithms," *Energy Build.*, vol. 239, p. 110839, 2021, doi: 10.1016/j.enbuild.2021.110839.

[7] S. Walker, W. Khan, K. Katic, W. Maassen, and W. Zeiler, "Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings," *Energy Build.*, vol. 209, p. 109705, 2020, doi: 10.1016/j.enbuild.2019.109705.

[8] J. Liu, H. Chen, J. Liu, Z. Li, R. Huang, L. Xing, J. Wang, and G. Li, "An energy performance evaluation methodology for individual office building with dynamic energy benchmarks using limited information," *Appl. Energy*, vol. 206, no. July, pp. 193–205, 2017, doi: 10.1016/j.apenergy.2017.08.153.

[9] C. V. Gallagher, K. Bruton, K. Leahy, and D. T. J. O'Sullivan, "The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings," *Energy Build.*, vol. 158, pp. 647–655, 2018, doi: 10.1016/j.enbuild.2017.10.041.

[10] V. J. Mawson and B. R. Hughes, "Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector," *Energy Build.*, vol. 217, p. 109966, 2020, doi: 10.1016/j.enbuild.2020.109966.

[11] S. Papadopoulos, E. Azar, W. L. Woon, and C. E. Kontokosta, "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation," *J. Build. Perform. Simul.*, vol. 11, no. 3, pp. 322–332, 2018, doi: 10.1080/19401493.2017.1354919.

[12] T. Ahmad and H. Zhang, "Novel deep supervised ML models with feature selection approach for large-scale utilities and buildings short and medium-term load requirement forecasts," *Energy*, vol. 209, p. 118477, 2020, doi: 10.1016/j.energy.2020.118477.

[13] S. Kumar, S. K. Pal, and R. P. Singh, "A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes," *Energy Build.*, vol. 176, pp. 275–286, 2018, doi: 10.1016/j.enbuild.2018.06.056.

[14] G. Ciulla, A. D'Amico, V. Lo Brano, and M. Traverso, "Application of optimized artificial intelligence algorithm to evaluate the heating energy demand of non-residential buildings at European level," *Energy*, vol. 176, pp. 380–391, 2019, doi: 10.1016/j.energy.2019.03.168.

[15] C. Miller, "What's in the box?! Towards explainable machine learning applied to non-residential building smart meter classificatio," *Energy Build.*, vol. 199, pp. 523–536, 2019, doi: 10.1016/j.enbuild.2019.07.019.

[16] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, "CASA, cost-optimal analysis by multi-objective optimisation and artificial neural networks: A new framework for the robust assessment of cost-optimal energy retrofit, feasible for any building," *Energy Build.*, vol. 146, pp. 200–219, 2017, doi: 10.1016/j.enbuild.2017.04.069.

[17] S. A. Sharif, A. Hammad, and P. Eshraghi, "Generation of whole building renovation scenarios using variational autoencoders," *Energy Build.*, vol. 230, p. 110520, 2021, doi: 10.1016/j.enbuild.2020.110520.

[18] S. Seyedzadeh, F. Pour Rahimian, S. Oliver, S. Rodriguez, and I. Glesk, "Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making," *Appl. Energy*, vol. 279, no. May, p. 115908, 2020, doi: 10.1016/j.apenergy.2020.115908.

[19] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, no. April 2017, pp. 1192–1205, 2018, doi: 10.1016/j.rser.2017.04.095.

[20] K. Koc, Ö. Ekmekcioğlu, and A. P. Gurgun, "Integrating feature engineering, genetic algorithm and tree-based machine learning methods to predict the post-accident disability status of construction workers," *Autom. Constr.*, vol. 131, no. August, 2021, doi: 10.1016/j.autcon.2021.103896.