

Traditional Machine Learning Models for Building Energy Performance Prediction: A Comparative Research

Zeyu Wu, Hongyang He*

College of Engineering and Physical Sciences, The University of Birmingham, Birmingham, United Kingdom

Email address:

Zxw285@student.bham.ac.uk (Zeyu Wu), HXH228@student.bham.ac.uk (Hongyang He)

*Corresponding author

To cite this article:

Zeyu Wu, Hongyang He. Traditional Machine Learning Models for Building Energy Performance Prediction: A Comparative Research. *Machine Learning Research*. Vol. 8, No. 1, 2023, pp. 1-8. doi: 10.11648/j.ml.20230801.11

Received: July 9, 2022; **Accepted:** July 27, 2022; **Published:** May 29, 2023

Abstract: A large proportion of total energy consumption is caused by buildings. Accurately predicting the heating and cooling demand of a building is crucial in the initial design phase in order to determine the most efficient solution from various designs. In this paper, in order to explore the effectiveness of basic machine learning algorithms to solve this problem, different machine learning models were used to estimate the heating and cooling loads of buildings, utilising data on the energy efficiency of buildings. Notably, this paper also discusses the performance of deep neural network prediction models and concludes that among traditional machine learning algorithms, GradientBoostingRegressor achieves better predictions, with Heating prediction reaching 0.998553 and Cooling prediction Compared with our machine learning algorithm HB-Regressor, the prediction accuracy of HB-Regressor is higher, reaching 0.998672 and 0.995153 respectively, but the fitting speed is not as fast as the GradientBoostingRegressor algorithm.

Keyword: Machine Learning, Building Energy, Decision Tree, Random Forest, Deep Learning, Gradient Descent Regression

1. Introduction

Predicting building energy consumption remains a challenging task because of the variety of factors that influence consumption, [1] such as the physical characteristics of the building, installed equipment, outdoor weather conditions and the energy use behaviour of the building's occupants [2, 3].

Physical modelling approaches and data-driven approaches are the two main methods used for predicting the energy consumption of buildings. For thorough energy modelling and analysis, physical models (sometimes called engineering techniques or white box models) rely on thermodynamic principles. EnergyPlus, eQuest, and Ecotect are a few examples of building energy modelling software that use physical models. [4] Based on specific architectural and environmental data, such as building construction specifications, operating schedules, HVAC design information, and climate, sky, and solar/shadow information, these kinds of software compute the energy consumption of buildings. At the time of the simulation, the user might not

have access to all of this specific information. Poor prediction performance may be the result of inaccurate inputs. [5]

There are numerous prediction techniques currently in use, including machine learning, and numerous papers have been released. There is, however, a dearth of research on the examination of machine learning and deep neural network methods side by side. To close this gap, we conduct a comprehensive comparison of deep learning and conventional machine learning methods in this study utilising actual and freely accessible datasets. [4] The results and future research objectives are reviewed. The assessment of prior research on energy prediction models for various building type parameters is the main objective.

The essay is set up like follows: The training and test datasets, which have undergone pre-processing, are initially displayed. This is followed by a theoretical overview of conventional machine learning algorithms, metrics to evaluate the algorithms' strengths and limitations, a comparative assessment of the prediction outcomes, and a look ahead to

Future development. [6]

2. Data Collection and Processing

The dataset used for the study in this paper was collected and processed by Athanasios Tsanas, a professor at the University of Oxford, UK, and consists of 12 different building shapes for energy analysis. These buildings differ in terms of parameters such as glazing area, glazing area distribution and orientation. Various settings were modelled as a function of the above characteristics to obtain 768 building shapes. The dataset contains 768 samples and 8

features, and we consider wall area, roof area, and glazing area as key indicators that can affect the energy load efficiency of both (heating and cooling) [5].

Information on the variables: including relative density surface area, wall area, roof area, total height, orientation, glazing area, glazing area distribution (variance) thermal load cooling load. [7] These eight characteristic factors do not have the same magnitude of influence on the energy performance of the building and the amount of data on these characteristic parameters varies slightly, as we can see from the following diagram which shows the table headings of the data set and the form of data collection.

Table 1. Headings of the data set.

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	Cooling Load
0	0.7638	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.9800	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.9800	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.9800	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.9000	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28

Filling of missing data values is done using the mean method of filling, where the impact weights of the eight indicators are assumed to be equal. The figure below shows the characteristics of the 8 main indicators, which will help the reader understand the training data and the test data. [9]

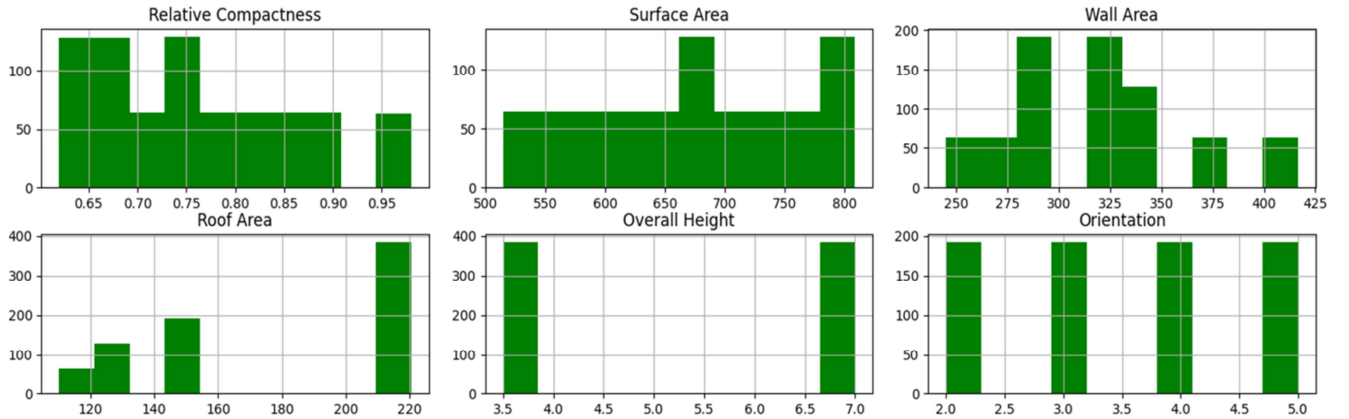


Figure 1. Six features, relative density surface area, wall area, roof area, total height, orientation.

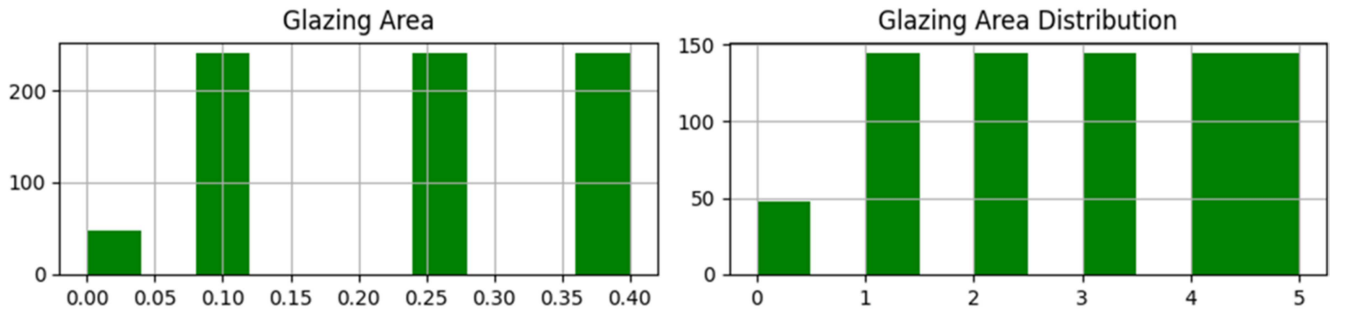


Figure 2. Two features, glazing area, glazing area distribution, Glazing area is a key indicator of the energy load that can be influenced.

The data is divided into training set, validation set and test set with k-fold cross-validation, and the ratio of training set, validation set and test set is 7:1:2. The method of dividing the data set is random sampling. The following figure shows the Correlation matrix of the training data.

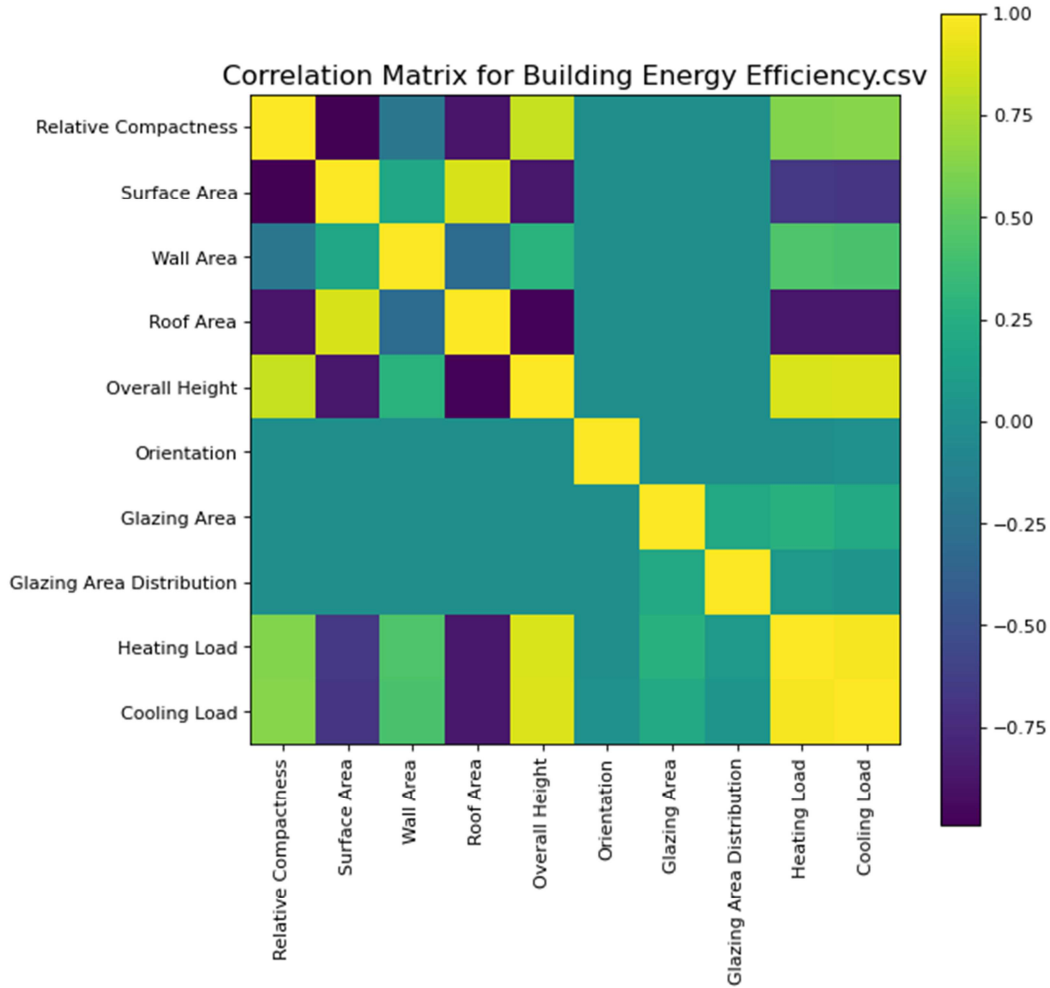


Figure 3. Correlation matrix of the training data.

3. Method

This chapter will use several types of machine learning algorithms that are currently in high use and have significant predictive efficiency, Decision Tree Algorithm, Random Forest Algorithm, Gradient Boosting Regression, all of which were significant at the time they were proposed and are based on data-driven predictive models. [10] In the field of building energy prediction, data collection involves collecting historical/available data for model training such as outdoor weather conditions and electricity consumption data. Data pre-processing can include data cleaning, data integration, data transformation and/or data reduction. Model training is the training of a model using a training dataset. Model testing is designed to evaluate the model using standard evaluation measures [8].

Decision trees have been a common approach to regression problems and the main focus of research in this area over the last few decades has been on the gradient boosting decision tree method (GBDT) based on the one proposed in. There are several open source packages that implement the GBDT algorithm (for both classification tasks and regression tasks). While the core ideas remain unchanged, these packages focus

mainly on speed-up, parallelisation, large-scale dataset processing and robust training, whereas this paper uses a decision tree regression model (DTR) and a grid search (GridSearchCV) to find the best combination of hyperparameters, GridsearchCV is a tuning process that uses hyper parameterisation to determine the optimal values for a given model. GridsearchCV performs an exhaustive search for a specified set of parameters. This method is computationally expensive, but produces good results [11, 12].

The decision tree consists of decision nodes, leaf nodes and the depth of the decision tree The decision tree is built based on information gain and the first step is to calculate the information entropy of the root node, which is calculated by the following formula:

$$H(D) = -\sum_{k=1}^N p_k \log_2 p_k \quad (1)$$

The information gain of the attributes is then computed, and the search for suitable attribute nodes is continued by repeating the previous approach. [11]

The algorithm is called the random forest algorithm because it combines multiple decision trees, with each dataset being selected at random with a random set of features selected as input. The algorithm first assumes that

the training set T is of size N , the number of features is M , and the size of the random forest is K . [10] The size of the random forest is traversed K times: a new sub-training set D is sampled N times from the training set T with put-back

sampling, and m features are randomly selected, where $m < M$. Using the new training set D and the m features, a complete decision tree is learned to obtain the random forest. [13, 14]

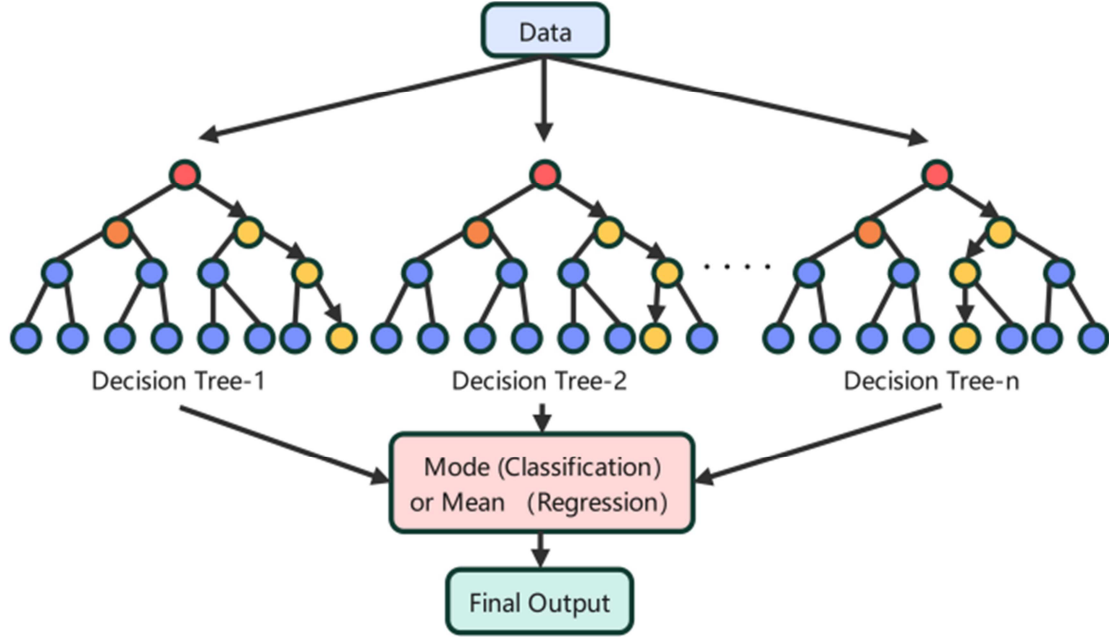


Figure 4. Correlation matrix of the training data.

Random forests reduce variance by averaging over many noisy but approximately unbiased trees, [15] thereby improving prediction accuracy. The variance of a random forest with a total number of trees (K) is

$$\rho\sigma^2 + \frac{1-\rho}{K}\sigma^2 \quad (2)$$

where σ^2 denotes the variance of individual trees, ρ denotes the correlation between trees, and M is the total number of trees in the set. Clearly, by increasing the total number of trees, M , the second term tends to zero. Thus, the variance of a random forest depends on three things [17, 18]:

The correlation ρ between any pair of trees: reducing the correlation reduces the total variance. This can be achieved by randomly selecting v from the p variables to split at each split node when growing trees on the bootstrap dataset. Reducing v reduces the correlation between trees and the strength of individual trees, and vice versa. Therefore, the optimal value of v needs to be found for a particular dataset. [16]

Variance per tree σ^2 , or in other words, strength per tree: strengthening the performance of each tree reduces the total variance of the model.

Total number of trees M : The second term of the equation can be reduced by increasing M . [19] Therefore, we should train a sufficient number of trees to ensure that the second term of the equation becomes zero.

In general, random forests are based on the idea of bagging, but with the diversity of each tree forced by random feature selection. The theoretical background of Random Forests supports parallel computing and thus its training can be

accelerated by parallel computing. The prediction performance of a random forest is influenced by three main factors: the correlation between individual trees, the performance of each tree and the total number of trees [20, 21].

Gradient boosting regression (GBR) is a technique that learns from its mistakes, and unlike bagging, the boosting method sequentially generates the underlying model. By focusing on these difficult-to-estimate training cases, prediction accuracy is improved by developing multiple models in sequence. During the boosting process, examples that are difficult to estimate using the previous base model appear more frequently in the training data than examples that are correctly estimated. Each additional base model is designed to correct the errors made by its previous base model. [22]

GBR proposes a modification to the gradient boosting method by using a fixed size regression tree as the base model. [23]

Assume that the number of leaves per tree is J . Each tree divides the input space into J disjoint regions $R_{1m}, R_{2m}, \dots, R_{jm}$ and predicts the constant value b_{jm} for region R_{jm} . The regression tree can be formally represented as

$$g_m(x) = \sum_{j=1}^J b_{jm} I(x \in R_{jm}) \quad (3)$$

Replacing $g_m(x_i)$ in the generic gradient boosting method with a regression tree, the model update equation and gradient descent step

$$f_m(x) = f_{m-1}(x) + \rho_m g_m(x) \quad (4)$$

$$\rho_m = \operatorname{argmin}_{\rho} \sum_{i=1}^n L(y_i, f_{m-1}(x_i) + \rho g_m(x_i)) \quad (5)$$

Becoming:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J \rho_m b_{jm} I(x \in R_{jm}) \quad (6)$$

$$\rho_m = \operatorname{argmin}_{\rho} \sum_{i=1}^n L(y_i, f_{m-1}(x_i) + \sum_{j=1}^J \rho b_{jm} I(x \in R_{jm})) \quad (7)$$

A separate best ρ_{jm} is used for each region R_{jm} and the b_{jm} can be discarded. The model update rule becomes:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J \rho_{jm} I(x \in R_{jm}) \quad (8)$$

$$\rho_{jm} = \operatorname{argmin}_{\rho} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \rho I(x \in R_{jm})) \quad (9)$$

4. Experimental Assessment Indicators

In order to compare the performance between algorithms, metrics are essential to assess the strengths and weaknesses of the models.

The mean absolute error refers to the average of the distance between the model prediction and the true value of the sample [21]. The formula for this is shown below:

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n} \sum_{i=1}^n |y_{\text{predict},i} - y_{\text{data},i}| \quad (10)$$

R^2 is the goodness of fit, which is the degree to which the regression line fits the observations. In statistics for line regression analysis of variables, when least squares is used for parameter estimation, R^2 is the ratio of the regression sum of squares to the total sum of squares of deviations [22], indicating the proportion of the total sum of squares of deviations that can be explained by the regression sum of squares, the larger this proportion the better. [23] The more accurate the model, the more significant the regression effect. r-squared is between 0 and 1, the closer to 1, the better the regression fit, and the better the fit is generally considered to be for models over 0.8.

$$R - \text{Squared } (R^2) = 1 - \frac{\sum_{i=1}^n (y_{\text{predict},i} - y_{\text{data},i})^2}{\sum_{i=1}^n (y_{\text{data},i} - \bar{y}_{\text{data}})^2} \quad (11)$$

One advantage of MAE over MSE is that MAE is less sensitive to predicting data outliers and is more inclusive:

$$\text{Mean Squared Error (MSE)} = \frac{1}{n} \sum_{i=1}^n (y_{\text{predict},i} - y_{\text{data},i})^2 \quad (12)$$

We generally use Accuracy and Error rate to evaluate the model as a whole, from a holistic perspective. [25]

$$\text{Error Rate } (\delta_i) = y_{\text{data},i} (1 - y_{\text{data},i}) (y_{\text{predict},i} - y_{\text{data},i}) \quad (13)$$

5. Experiments and Results

The experimental environment was based on the Pytorch 11.0 framework, CUDA 11.3 and CUDNN 8.2, and the training model was based on an NVIDIA GeForce RTX 3060 (12 GB). For DecisionTreeRegressor prediction, the

minimum sampling interval `min_samples_split` was initialized to 15, the maximum depth to 6, the maximum number of leaf ends to 31, and the minimum number of sampled leaves to 6. The optimal hyperparameters were determined by GridsearchCV with 'max_depth' of 6, 'max_leaf_nodes' of 31, 'min_samples_leaf' of 5, and 'min_samples_split' of 17 [26].

R^2 is an indispensable metric when measuring the fit of DecisionTreeRegressor, Random Forests and Gradient Boosting Regression. The DTR algorithm has an R-Squared of 0.991543907727448 for the training dataset and an R-Squared for the test data set is 0.973711991905928. The random forest algorithm initialises the hyperparameters 'n_estimators' at [350, 400, 450], 'max_features' at [1, 2], 'max_depth' at [85, 90, 95], and the best hyperparameters obtained by the CV algorithm are 'max_depth' is 90, 'max_features' is 1, 'n_estimators' is 450, and it is worth noting that the R-Squared for the training dataset of the Random Forest algorithm is 0.991543907727448 The R-Squared for the test dataset is 0.973711991905928. Gradient Boosting Regression has one more hyperparameter subsample than the above two algorithms. In this experiment, we set the downsampling to 1.0 and the R-Squared of the training dataset is 0.998672666920205.

The R-Squared of the test dataset is 0.9914370717646062.

From the data representation of R-Squared alone the prediction fit of Gradient Boosting Regression is the best, HB-Regressor has 6 hidden layers, of which the first layer has 180 neurons as input and the activation function is RULE, the RULE function can well eliminate the gradient disappearance problem: ReLU has a constant gradient of 0 for inputs greater than The gradient of ReLU is constant at 1 when the input is greater than 0, thus avoiding the training difficulties caused by the gradient disappearance problem, and the number of iterations set is 10000 [26].

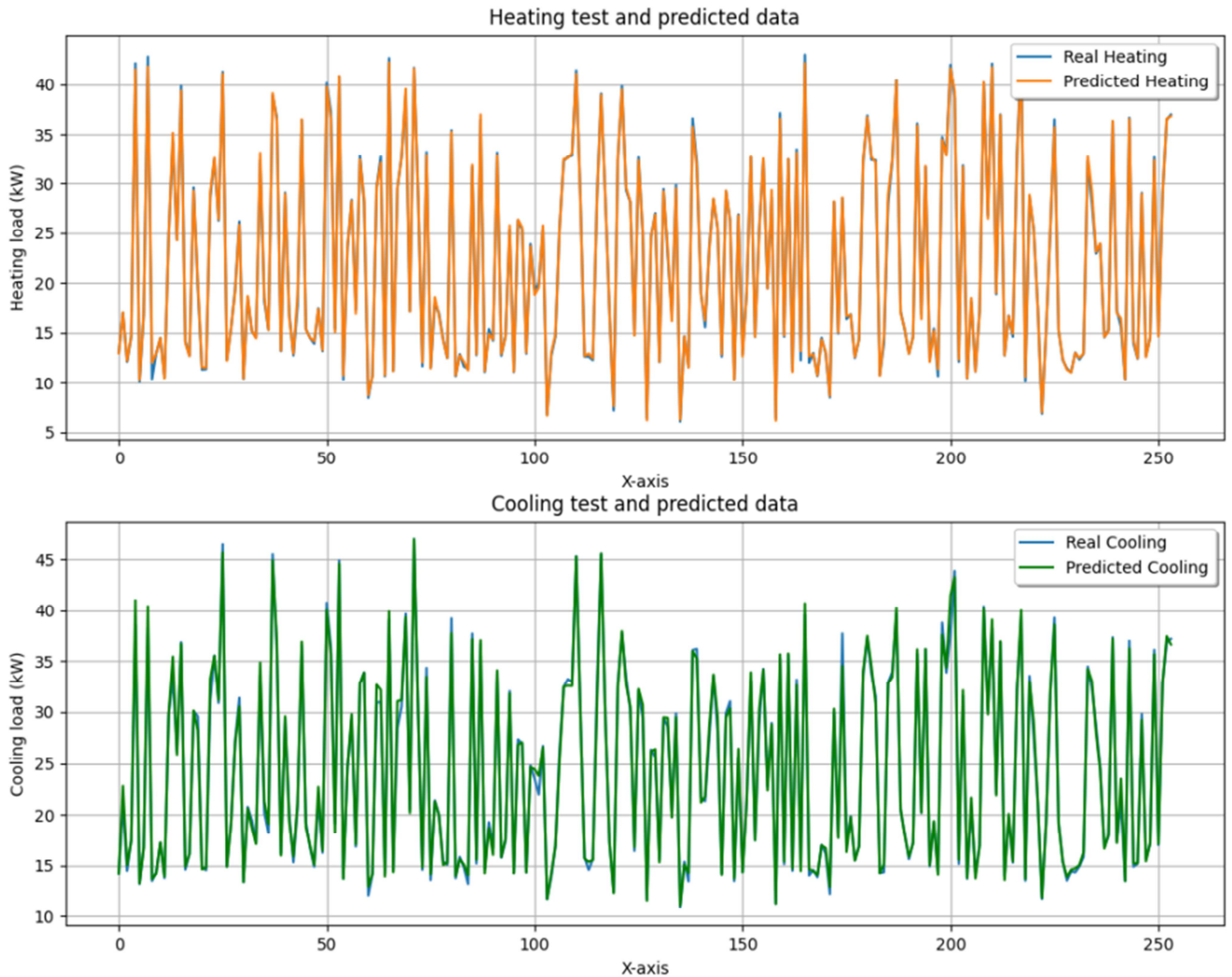
The table below shows the prediction accuracy from the heating training set, the prediction accuracy from the heating test set, the prediction accuracy from the cooling training set, the prediction accuracy from the cooling test set and the R-Squared index.

Table 2. Algorithm comparison.

	Model	Train_Heating	Test_Heating	Train_Cooling	Test_Cooling	R ²
0	DecisionTreeRegressor	0.994803	0.995168	0.967655	0.959112	0.973711
1	RandomForestRegressor	0.998695	0.991333	0.996960	0.973201	0.991543
2	GradientBoostingRegressor	0.999735	0.998553	0.998988	0.991905	0.991437
3	HB-Regressor	0.999821	0.998672	0.997545	0.995153	0.994384

We can see that the prediction accuracy of HB-Regressor on the heating training set, and the prediction accuracy on the heating test set are slightly higher than the other three machine learning algorithms, but the prediction accuracy on the test set is 99.8672% slightly lower than the accuracy on the training set, and the fit of HB-Regressor is the most excellent, reaching 99.4384%, and the HB-Regressor is excellent overall, but the prediction accuracy on the cooling training dataset is slightly worse than that of the GBR algorithm.

The graph below shows the predicted versus true value curves for the cooling and heating data for HB-Regressor. [24] The predicted and true value curves are coloured differently and we can see that the predicted values are correct in most cases. As the forecasting model calculates the heating demand based on the input of the data time, the forecast range becomes a key parameter for the accuracy of the estimation using measured or forecast data.

*Figure 5. Heating test and predicted data/cooling test and predicted data.*

6. Discussion

Based on the analysis of the case studies in this research and the methodologies used there, machine learning deep network

prediction methods outperformed techniques like decision trees, random forests, etc. in terms of accuracy. This could be the outcome of the input selection procedure, which disregards the historical lags of the input variables. The use of such lags may be advantageous for prediction techniques, which

frequently rely on recent and historical data. Real-time access to the measured data is necessary in this situation. Although predictive methods perform better, they have the drawback of requiring a lot of data to support them, much of which may not be readily accessible in real life. Prediction techniques that rely on the current time step value nonetheless perform well enough with such an input set of data [18].

All ML models seem to have acceptable target variable errors on the test data when focusing on prediction error. Given each application, the deep learning and updated tree-based models seem to be close to the low end of the error range. However, one of the best performing algorithms in this study was also discovered to be the HB-Regressor. The case study may not have explored the latency of the inputs and/or the potential length of the dataset, which may have better highlighted the advantages of the deep learning approach. As a result, the differences seen across the test dataset appear to be small.

Since the 1990s, numerous studies have created various forecasting models for energy use in buildings. Physical approaches and data-driven methods are the two basic categories into which these techniques fall. The physical technique, which is considered the classic method used in the design phase for building energy evaluation, is primarily based on the use of energy modelling tools. Support vector machines (SVM), random forests (RF), and other machine learning (ML) techniques, which need fewer construction parameters, are frequently used in data-driven methodologies. Furthermore, because findings can be obtained quickly, data-driven solutions have been found to be more accurate and efficient. Decision trees (DT), artificial neural networks (ANN), and support vector machines (SVM), to mention a few, were some of the most efficient techniques used. These algorithms have all occasionally performed better than one another. It was done to compare various models, including the multilayer perceptron ANN and SVM. It was determined that ANNs perform better in predicting energy use than SVMs. Additionally, feature selection and hyperparameter tuning are significant factors that frequently have an impact on the performance of ML models. The model's findings are significantly impacted by this.

7. Future Work

Future research work could also explore the application of other types of machine learning algorithms in energy consumption prediction. For example, deep learning algorithms have been shown to outperform other machine learning algorithms in many other areas (e.g. image classification and multimodal data analysis), but have not been fully investigated in the area of building energy consumption prediction.

As new data-driven models are developed, sharing more information about the development process and purpose, validation and reusability of these models is essential to avoid unnecessary duplication of research efforts. Some

important model information (for example, the purpose of the prediction) is sometimes not reported or not adequately described. Insufficient information provides limited guidance on the applicability of certain models to new contexts, which may inhibit model reusability [27].

8. Conclusion

It is obvious that a high performance ML model can be developed to forecast building energy use at the design stage given the high accuracy in this study. The ability to estimate building energy use has been shown through previous research. There isn't a recognised top ML model during the design phase, though. As they can provide energy performance results in just a few seconds, machine learning models are far more effective than conventional simulation techniques. This has been proven in numerous earlier investigations. However, no research has been done to create models for the building design phase by implementing feature selection methods and hyperparameter tuning a variety of algorithms. Accuracy and R Square measures were used to assess four base models.

In terms of predicting building energy performance, Gradient Boosting (GB) fared better than other conventional machine learning models. This study compared different machine learning models for predicting energy performance, and the results showed that integrated deep learning models outperformed other models that could support their liability to produce high-quality results because they were trained on a variety of algorithms that were theoretically similar to the theory. These results, however, go against some previous studies where the GBR algorithm fit produced superior prediction accuracy at lower levels. The findings demonstrated that the GBR algorithms lacked both performance and computational efficiency. However, this is explained by the idea that GBR algorithms perform better in small datasets whereas the deep learning algorithm HB performs better in large datasets.

Abbreviations

The following abbreviations are used in this manuscript:

GBR: structural health monitoring

RF: Random Forest

DT: Decision Tree

CNN: Convolutional Neural Network

DNN: Deep Neural Network

GBDT: gradient boosting decision tree

MAE: Mean Absolute Error

MSE: Mean Squared Error

Author Contributions

All authors have contributed toward this work as well as in compilation of this manuscript. All authors read and approved the final manuscript.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Please contact author for Supplementary Materials.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Amasyali, Kadir, and Nora M. El-Gohary. "A review of data-driven building energy consumption prediction studies." *Renewable and Sustainable Energy Reviews* 81 (2018): 1192-1205.
- [2] Li, Chengdong, et al. "Building energy consumption prediction: An extreme deep learning approach." *Energies* 10.10 (2017): 1525.
- [3] Qiao, Qingyao, Akilu Yunusa-Kaltungo, and Rodger E. Edwards. "Towards developing a systematic knowledge trend for building energy consumption prediction." *Journal of Building Engineering* 35 (2021): 101967.
- [4] Wang, Ran, Shilei Lu, and Wei Feng. "A novel improved model for building energy consumption prediction based on model integration." *Applied Energy* 262 (2020): 114561.
- [5] Liu, Zhijian, et al. "Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction." *Energy Exploration & Exploitation* 37.4 (2019): 1426-1451.
- [6] Dong, Zhenxiang, et al. "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification." *Energy and Buildings* 241 (2021): 110929.
- [7] Lei, Lei, et al. "A building energy consumption prediction model based on rough set theory and deep learning algorithms." *Energy and Buildings* 240 (2021): 110886.
- [8] Rigatti, Steven J. "Random forest." *Journal of Insurance Medicine* 47.1 (2017): 31-39.
- [9] Kotsiantis, Sotiris B. "Decision trees: a recent overview." *Artificial Intelligence Review* 39 (2013): 261-283.
- [10] Zhang, Yanru, and Ali Haghani. "A gradient boosting method to improve travel time prediction." *Transportation Research Part C: Emerging Technologies* 58 (2015): 308-324.
- [11] Nie, Peng, et al. "Prediction of home energy consumption based on gradient boosting regression tree." *Energy Reports* 7 (2021): 1246-1255.
- [12] He, Hongyang. "Research on the application of electronic technology of internet of things in smart city." 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, 2020.
- [13] Jetcheva, Jorjeta G., Mostafa Majidpour, and Wei-Peng Chen. "Neural network model ensembles for building-level electricity load forecasts." *Energy and Buildings* 84 (2014): 214-223.
- [14] Catalina, Tiberiu, Vlad Iordache, and Bogdan Caracaleanu. "Multiple regression model for fast prediction of the heating energy demand." *Energy and buildings* 57 (2013): 302-312.
- [15] Yun, Kyungtae, et al. "Building hourly thermal load prediction using an indexed ARX model." *Energy and Buildings* 54 (2012): 225-233.
- [16] Popescu, D., and F. Ungureanu. "Prediction of space heating consumption in district heated apartments." *ASME International Mechanical Engineering Congress and Exposition*. Vol. 56291. American Society of Mechanical Engineers, 2013.
- [17] Wong, Sai Li, Kevin KW Wan, and Tony NT Lam. "Artificial neural networks for energy analysis of office buildings with daylighting." *Applied Energy* 87.2 (2010): 551-557.
- [18] Kamaev, V. A., et al. "Using connectionist systems for electric energy consumption forecasting in shopping centers." *Automation and Remote Control* 73.6 (2012): 1075-1084.
- [19] Wu, Xindong, et al. "Top 10 algorithms in data mining." *Knowledge and information systems* 14 (2008): 1-37.
- [20] Fumo, Nelson. "A review on the basics of building energy estimation." *Renewable and sustainable energy reviews* 31 (2014): 53-60.
- [21] Deák, István. "Random number generators and simulation." *Mathematical methods of operation research* (1990).
- [22] Kwok, Simon SK, and Eric WM Lee. "A study of the importance of occupancy to building cooling load in prediction by intelligent approach." *Energy Conversion and Management* 52.7 (2011): 2555-2564.
- [23] Becerik-Gerber, Burcin, et al. "Civil engineering grand challenges: Opportunities for data sensing, information analysis, and knowledge discovery." *Journal of Computing in Civil Engineering* 28.4 (2014): 04014013.
- [24] Jain, R. K., T. Damoulas, and C. E. Kontokosta. "Towards data-driven energy consumption forecasting of multi-family residential buildings: feature selection via the lasso." *Computing in Civil and Building Engineering* (2014). 2014. 1675-1682.
- [25] Fan, Cheng, Fu Xiao, and Shengwei Wang. "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques." *Applied Energy* 127 (2014): 1-10.
- [26] Chou, Jui-Sheng, and Dac-Khuong Bui. "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design." *Energy and Buildings* 82 (2014): 437-446.
- [27] Li, Qiong, Peng Ren, and Qinglin Meng. "Prediction model of annual energy consumption of residential buildings." 2010 international conference on advances in energy engineering. IEEE, 2010.