2024 European Conference on Computing in Construction
Chania, Crete, Greece
July 14-17, 2024

EVALUATING THE CAPABILITIES OF SURROGATE MODELING TECHNIQUES IN PREDICTING HOURLY BUILDING ENERGY CONSUMPTION

Shivram Krishnan[1], Anagha Divakaran[1], Omprakash Ramalingam Rethnam[1], Elin Markarian[2], Seif Qiblawi[2], Albert Thomas[1], and Elie Azar[2]
[1] Indian Institute of Technology Bombay, Mumbai, India
[2] Carleton University, Ottawa, Canada
Corresponding author, e-mail address: albert@iitb.ac.in

Abstract

Building energy simulation models have strong energy prediction capabilities but suffer from high computational costs, which could be reduced through surrogate modeling approaches. Existing surrogate models predict energy consumption on an annual resolution, however, for strategizing net-zero measures, granular predictions are necessary. This paper evaluates the ability of four state-of-the-art machine learning algorithms to predict building energy consumption on an hourly basis. The results indicate that Random Forest Regression is the most suitable predictive model due to the high R^2 value of 0.94. The proposed framework can be further expanded to test net-zero energy retrofits at minimal computational costs.

Introduction

The operation phase of buildings accounts for nearly 30% of global emissions (IEA 2021) which necessitates the need for innovative strategies for reducing emissions. The International Energy Agency estimates that in the next 30 years, global building floor area will grow by 75%, with around 80% of the increase in emerging markets such as India. The building sector in India accounts for over 30% of total electricity consumed (BEE 2018) and the commercial building sector in India consumed 8.31% of end-use electricity in the year 2021 (MoSPI 2022). Previous studies on energy consumption show that the availability of an accurate building energy forecasting system can save about 30% of the total energy consumption in buildings (Olu-Ajayi et al. 2022). Hence, predicting the energy consumption pattern of buildings is of significance to multiple stakeholders as it can help improve buildings' energy efficiency and aid in energy conservation measures for buildings.

Building energy simulations for energy prediction are typically done by creating a prototype building model using an energy modeling tool such as EnergyPlus, developed by the U.S. Department of Energy based on the proposed design inputs (Chan et al. 2022). Data-driven approaches using surrogate machine learning (ML) based models help overcome the challenges of building an energy simulation approach, which requires high computational costs for performing the energy simulations. The ML models identify the relationship between input features impacting building performance such as weather parameters and output variables such as energy consumption. Such a trained ML model develops the capability to predict the energy consumption of buildings eliminating the need for multiple energy simulation runs for varying simulation scenarios (Olu-Ajayi et al. 2022).

The development of a surrogate ML model consists primarily of four steps: data collection, data preprocessing, model training, and model testing. Data collection involves collecting available past data, such as weather data and energy consumption data of buildings. Data preprocessing is done by integrating the required data into a comprehensive dataset, which will be used further for model training. The model training step involves developing and deploying a suitable ML model for energy prediction studies. Model testing involves using standard model evaluation measures to check the accuracy of the prediction model (Amasyali & El-Gohary, 2018).

Energy prediction using machine learning models is commonly done using algorithms such as regression models and artificial neural networks (ANN). Some of the commonly employed regression models in energy prediction include multiple linear regression (MLR), polynomial regression (PR), and random forest regression (RFR) (Sun et al. 2020). In a study based out of buildings in Italy, multiple linear regression is used as a decision support tool for assessing the preliminary energy demand. This was used as an alternative to assessing energy demand by building energy modeling. The correlation of the results shows that the MLR model can predict the buildings' heating, cooling, and total energy demand with a high degree of reliability (Ciulla et al. 2019). Meanwhile, the polynomial regression model is more flexible than multiple linear models as it can fit data in a wide range of curvature, making it suitable for problems where the relationship between input and output variables is non-linear, as in the case of building energy loads. In a study conducted on 17 buildings in Europe, polynomial regression was used to predict the heating demand based on climatological conditions and the architecture of the building. The results showed that the polynomial regression model prediction is best suited for predicting fast, early-stage energy consumption (Tiberiu 2012).

Another ML model that can capture non-linear relationships between input and target variables is Random Forest Regression. In a study conducted on a dataset of 5 existing buildings from a building data genome project (Miller et al. 2017), short-term energy consumption in hourly resolution is predicted to propose effective solutions to help buildings' owners and facility managers understand building energy consumption patterns for enhancing energy efficiency in buildings. The RFR model was selected for the study as it can reduce predictive error while solving regression problems, and it considers all decision trees in the prediction correlation (Pham et al. 2020). The results from the study showed that RFR is an effective technique in predicting hourly building energy consumption, which helps stakeholders take adequate steps to reduce the energy consumption of buildings.

Artificial neural network (ANN) is also a commonly used ML model for predicting the energy consumption of buildings besides RFR. In a study of 243 commercial buildings in Southeast Asia, the cooling loads of buildings obtained from the physics-based building energy model were compared to an ANN model. ANN was used due to its ability to solve non-linear problems dependent on numerous variables. The features for developing the ANN model were general building information, building envelope, and internal loads. The results showed that the cooling load prediction by the ANN model agrees well with the physics-based building energy simulation at a fraction of the time required for the latter (Ngo 2019).

Overall, existing studies have largely focused on predicting the 'annual' energy consumption of buildings, and energy prediction for 'hourly' resolution and 'one-year' time horizon has not been addressed thoroughly. The current studies are also limited to using meteorological information from a single climatic zone, and the impact of variation of varied climatic conditions on energy consumption is not explored in detail. Hence, this study aims to address this critical gap by identifying the best ML model for predicting building energy consumption in hourly resolution under varying climatic conditions for an annual time horizon. The developed framework is expected to reduce the computational effort required by a traditional building energy modeling approach for hourly energy predictions, even under varying climatic conditions, by eliminating the need for multiple energy simulation runs. This approach can further be used in optimization studies to inform design strategies, such as renewable energy generation, that support the transition to net-zero energy performance for existing buildings.

Methodology and Experiment

A four-stage methodology to identify the best predicting ML model for hourly energy consumption is proposed as follows:

Stage 1: Generating hourly energy simulation using Energy Plus

Stage 2: Compiling dataset for ML training

Stage 3: Developing ML Model and evaluating

Stage 4: Predicting energy consumption using ML model

The flow chart of the hourly energy prediction framework is given in Figure 1 below.

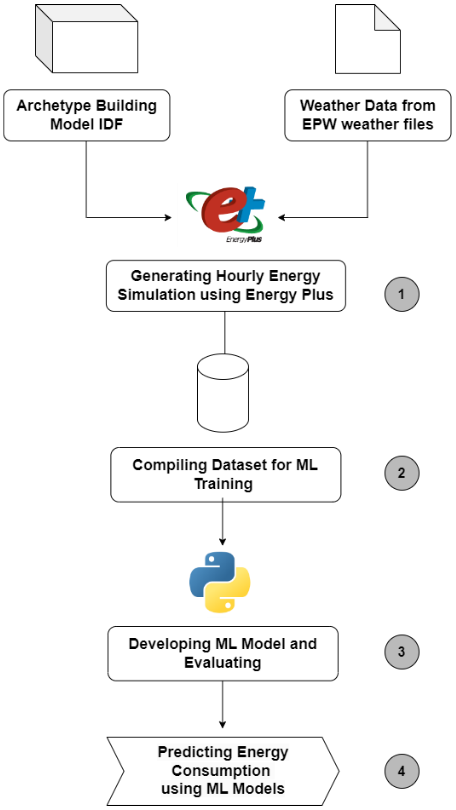

Figure 1 Framework for hourly energy prediction

Stage 1: Generating Hourly Energy Simulation Using Energy Plus

The dataset used for ML training is the collection of meteorological data from weather files and hourly building energy consumption data generated from energy simulations. The process is explained in detail below.

Hourly building energy consumption data based on annual simulation of a standard reference building model is used for the ML model development. A validated prototype reference commercial building model, based on a previous study of 230 commercial buildings in India (Bhatnagar et al. 2019), is adopted to simulate hourly energy consumption, and the archetype-building model used for this study is shown in Figure 2 below.

Archetype Building Model

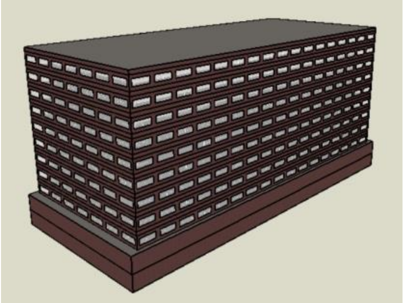

Figure 2 Archetype reference commercial building model for an 8-hour working schedule

Building features of the archetype reference building used in this experiment are mentioned in Table 1 below.

Table 1 Archetype reference building features

Feature	Value
Total floor area (m2) (excluding basement)	31,381
Percent of conditioned Area (%)	76
Basement floor area (m2)	7,689
Floor Area	3,487
Number of floors	9
Building shape	Rectangular
Aspect ratio	2.3
Window-to-wall ratio	30%
Window U-factor (W/m2-K)	2.05
Window SHGC	0.27
Exterior wall	Brick wall (200mm width) with plaster
Exterior wall U-factor	1.46
Roof	Roof tile + Concrete (200 mm) + Expanded polystyrene (50mm) + Plaster
Roof U-factor	0.64
Average lighting power density (W/m2)	Office: 8.32; Basement parking: 1.81
Daylighting controls	No
Occupancy controls	No
Exterior Light (kW)	8.12
Cooling type	Screw Chiller
Distribution and terminal	Variable air volume AHU with motorized dampers
COP	5.6
Supply Fan Fan Power (W/l/s)	1.25
Economizers	No
Energy recovery	No

Weather data

The weather data in 'epw' format, developed by the U.S. Department of Energy, is used for this study. Representative cities used for this study are shown in Table 2. The input variables (X-variables) used for ML model development, are listed in Table 3. The weather file input parameters are also taken as per studies conducted previously. In addition to the weather parameters, occupancy fraction is also considered as one of the input variables for energy consumption prediction as the occupancy schedule has been shown to significantly influence energy consumption (Li et al. 2018).

Table 2 Representative cities for ML model development

Sl No:	Climatic Zone	Representative City
1	Hot-Dry	Ahmedabad
2	Hot-Humid	Kolkata
3	Temperate	Bengaluru
4	Composite	Lucknow

Table 3 List of input variables for ML model development

Sl.No.	Input Variables (X)	Variable Name
1	Dry bulb temperature	X1
2	Dew point temperature	X2
3	Relative humidity	X3
4	Global horizontal radiation	X4
5	Direct normal radiation	X5
6	Diffuse horizontal radiation	X6
7	Wind direction	X7
8	Wind speed	X8
9	Occupancy	X9

Stage 2: Dataset for ML Training

The simulation software 'EnergyPlus,' version 8.7, developed by the U.S. Department of Energy, is used to generate hourly energy consumption for an annual simulation with 8,760 hourly energy data points. Out of the total 8,760 data, an 8-hour working schedule, from morning 09:00 to evening 17:00, is considered. The days considered were from Monday to Friday, excluding 10 public holidays. The resulting dataset has 2,000 hourly data points of input (X-variables) and output variables (Y-variable) as given in Table 4.

Table 4 Data points considered based on an 8-hr schedule

Climate Zone	City Name	Total Simulation Data	Data points based on 8-hr schedule
Hot-Dry	Ahmedabad	8,760	2,000
Hot-Humid	Kolkata	8,760	2,000
Temperate	Bengaluru	8,760	2,000
Composite	Lucknow	8,760	2,000

Stage 3: ML Model Development and Evaluation

The dataset is further split into training and test sets. The 'X-variables' and 'Y-variables' from the dataset are split in an 80:20 ratio for all the subsequent ML models used to develop a predictive model. Four different ML algorithms, such as MLR, PR, RFR, and ANN, are used to investigate the best-performing model in predicting hourly energy consumption. MLR, PR, and RFR algorithms are developed using the 'scikit-learn' library, and the ANN model is developed using the 'tensorflow' library with Python as the programming language. ML model development is given in more detail in the subsequent sections.

Multiple Linear Regression

From the 'scikit-learn' Python library, the 'LinearRegression' function is imported. The split dataset is further given as input to the linear regression function. The 'predict' function is then used with the test set of 'X-variables' to predict the hourly energy consumption.

Polynomial Regression

Similar to the multiple linear regression, 'PolynomialFeatures' and 'LinearRegression' functions are imported from the 'scikit-learn' library. The training dataset for 'X-variable' is then transformed using 'fit_transform'. The 'predict' function is further used with the test set of 'X-variables' to predict the hourly energy consumption.

Random Forest Regression

For RFR, the 'RandomForestRegression' function is imported from the 'ensemble' module from the 'scikit-learn' library. The 'RandomForestRegression' function is then called on the split dataset with 'n_estimators' defined as 10 and 'random_state' set as 0. The 'predict' function is further used with the test set of 'X-variables' to predict the hourly energy consumption.

Artificial Neural Network

ANN model differs slightly from the other three models used. For ANN, the 'Tensorflow' python library is used for analysis. A multi-layered feed-forward ANN model with one input layer, one hidden layer, and one output layer is used based on trial and error for prediction accuracy. The number of 'epochs' is set as 100. The 'predict' function is further used with the test set of 'X-variables' to predict the hourly energy consumption.

Following the ML model development, model performance evaluation techniques are further deployed on each ML model to select the best predicting model. The model performance evaluation is conducted separately for the four representative cities.

Stage 4: Energy Prediction using ML Model

Some of the standard model performance evaluation metrics used for evaluating the test set prediction results are coefficient of determination (R^2), Mean Squared Error (MSE), Mean Bias Error (MBE), Root Mean Squared Error (RMSE), and coefficient of variance (CV) (Sun 2020). All these model performance parameters are estimated based on the equations below, and the results are discussed in further sections.

The coefficient of determination (R^2), as mentioned in Equation 1, measures how close the predicted value is to the actual value, with a value of 1 indicating a perfect fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(PEi - AEi)^2}{\sum_{i=1}^{n}(AEi - \overline{AEi})^2} \qquad (1)$$

The Mean Squared Error (MSE), mentioned in Equation 2, calculates the variance between actual and predicted values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(AEi - PEi)^2 \qquad (2)$$

The Root Mean Squared Error (RMSE), in Equation 3, represents the square root of the quadratic mean of the differences between predicted and expected values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AEi - PEi)^2} \qquad (3)$$

The Mean Bias Error (MBE), mentioned in Equation 4, is the average error or bias of each value between the actual and predicted values in the sample space.

$$MBE = \sum_{i=1}^{n}\frac{(AEi - PEi)}{n} \qquad (4)$$

The Coefficient of Variation (CV), mentioned in equation 5, indicates how well a model approximates the real data points, which measures the model's degree of predictability.

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^{n}(PEi - AEi)^2}{n-1}}}{\overline{AEi}} * 100 \qquad (5)$$

where PEi is the predicted energy, AEi is the actual energy at the hour 'i', $\overline{AEi}$ is the mean of actual energy, and n is the total number of samples.

Based on the results obtained from the model performance evaluation conducted for each representative city, the best predicting ML model is selected for further scenario analysis.

Scenarios considered using the best predicting ML Model

The best ML algorithm, selected based on model performance evaluation, is further deployed in two different scenarios. The two different scenarios considered in the study are explained below.

Scenario 1

In Scenario 1, the X and Y variables from all four representative cities are considered and combined to train the ML model. The combined dataset, which contains 8,000 data points, is then split into an 80:20 ratio. The hourly energy consumption of the building using the test set is then predicted.

Scenario 2

Scenario 2 demonstrates the energy forecasting capacity of the ML model for a dataset it is not trained with. In this case, the data points from three climatic zones were used, and the combined dataset contained 6,000 data points. In this study, the training set included X and Y variables of Ahmedabad, Lucknow, and Bengaluru, and the hourly energy consumption of Kolkata is predicted.

Results and Discussion

The experiment to predict the hourly energy consumption using ML models was conducted and the results are given and discussed as follows.

Model Performance Evaluation

The results of the model performance evaluation conducted for the four ML models MLR, PR, ANN, and RFR using Lucknow's dataset are shown in Figure 3 and Table 5 below. Similarly, the model performance evaluation is conducted for the Kolkata, Ahmedabad, and Bengaluru datasets, and the results are shown in Table 6 to Table 8, respectively.

Table 5 Model performance evaluation for Lucknow with MLR, PR, ANN, and RFR models

Model performance	**R^2**	**MSE**	**RMSE (%)**	**MBE (%)**	**CV (%)**
MLR	0.89	1169.6	34.20	-1.27	9.6
PR	0.83	1857.6	43.10	-4.58	10.9
ANN	0.88	1296.2	36.00	-3.20	10.1
RFR	0.94	636.39	25.22	-2.88	9.7

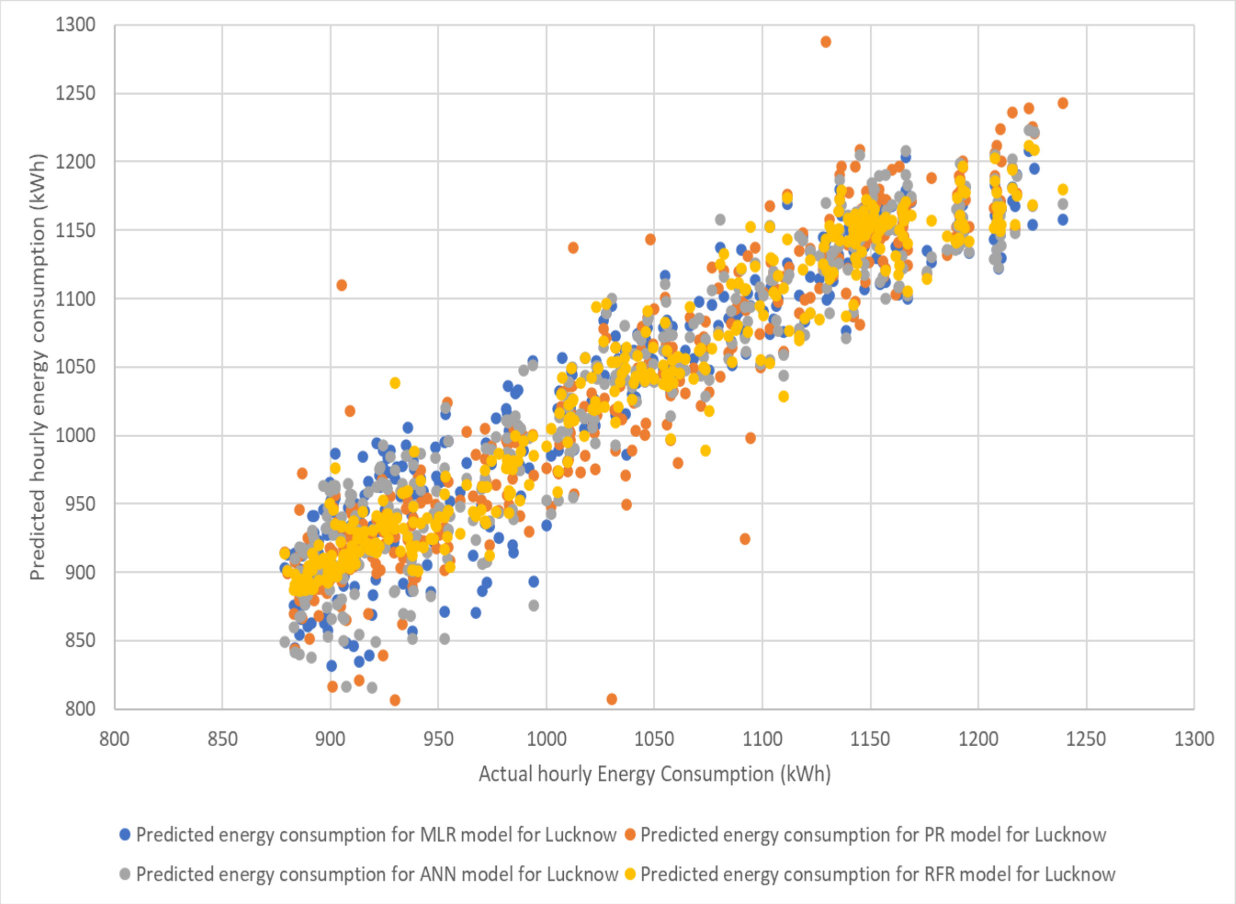

Figure 3: Graph showing the predicted vs actual hourly energy consumption for the four ML models tested for Lucknow

Table 6 Model performance evaluation for Kolkata with MLR, PR, ANN, and RFR models

Model performance	**R^2**	**MSE**	**RMSE (%)**	**MBE (%)**	**CV (%)**
MLR	0.86	1208.1	34.75	1.72	8.9
PR	0.56	4055.1	63.67	4.31	10.4
ANN	0.82	1574.4	39.67	7.54	9.7
RFR	0.91	748.4	27.37	1.55	8.8

Table 7 Model performance evaluation for Ahmedabad with MLR, PR, ANN, and RFR models

Model performance	**R^2**	**MSE**	**RMSE (%)**	**MBE (%)**	**CV (%)**
MLR	0.75	1637.88	40.47	5.80	7.0
PR	0.42	3846.89	62.02	-1.02	9.3
ANN	0.75	1667.55	40.83	-5.89	8.0
RFR	0.88	739.46	27.19	1.69	7.4

Table 8 Model performance evaluation for Bengaluru with MLR, PR, ANN, and RFR models

Model performance	R^2	MSE	RMSE (%)	MBE (%)	CV (%)
MLR	0.61	559.18	23.64	1.110	3.4
PR	0.25	1071.9	32.73	-1.105	4.6
ANN	0.50	709.73	26.64	-3.126	3.9
RFR	0.70	434.28	20.83	1.904	3.4

The graphs shown in Figure 3, represent predicted and actual energy consumption for the four ML models MLR, PR, ANN, and RFR for the city of Lucknow. The X-axis represents actual hourly energy consumption and the Y-axis represents predicted hourly energy consumption in kWh. In the case of the RFR model, in Figure 3, the predicted energy consumption closely follows the actual energy consumption. However, in the case of MLR, PR, and ANN, the predicted energy consumption shows variations. In some instances of MLR, PR, and ANN, the predicted energy consumption value is higher than actual energy consumption. Thus, the RFR model shows better performance when compared to MLR, PR, and ANN.

The results observed from Table 5 to Table 8 show that the RFR model outperforms other models due to its high R^2 values and lower MSE, RMSE MBE, and CV values. Hence, the RFR model is chosen as the prediction model for further scenarios 1 and 2.

Energy Prediction Results for Scenario 1 and 2

Scenario 1

Hourly energy consumption is predicted using the RFR model, trained with the dataset of all four representative cities combined. The results are shown in Table 9 below.

Table 9 RFR model performance in the test phase of scenario 1

Scenario	Training points	R^2	MSE	RMSE (%)	MBE (%)	CV (%)
I	6,400	0.90	837.8	28.94	2.21	8.70

The performance evaluation of the test data in the test phase of scenario 1 gives an R^2 value of 0.90 and a CV value of 8.70%, indicating that the RFR model is reliable for hourly building energy prediction (Yang et al. 2015).

Further, to understand the results figuratively, Figure 4 shows the predicted versus actual hourly energy consumption for the test data of scenario 1. The scatter plot is generated from the test set of scenario 1, i.e., with 1600 data points (20% of 8000 data points) combined from all four cities. The concentration of the scatter plot points along the 45-degree line indicates a high degree of correlation between actual and predicted values.

Further to test the hourly energy prediction capability of the RFR model with changing weather conditions scenario 2 was conducted and the results are given in the following section.

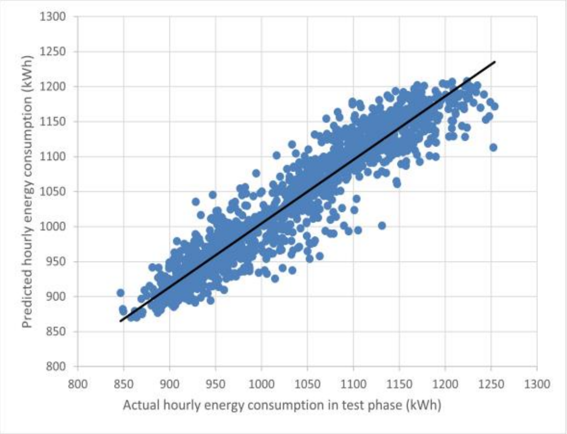

Figure 4: Scatter plot of predicted vs actual hourly energy consumption in the test phase of scenario 1

Scenario 2

Hourly energy consumption is predicted using the RFR model trained with a dataset of three cities: Ahmedabad, Bengaluru, and Lucknow, and the energy consumption of Kolkata is predicted. The performance of the RFR model is given in Table 10.

Table 10 RFR model performance in test phase of scenario 2

Scenario	Training points	R^2	MSE	RMSE (%)	MBE (%)	CV (%)
II	4,800	0.80	1825.1	42.72	8.96	8.80

The model performance evaluation shows an R^2 value of 0.80 and a CV of 8.80%. The results indicate that the prediction accuracy of the RFR model for scenario 2 is less than that for scenario 1. To understand the results figuratively, Figure 5 shows the predicted versus actual hourly energy consumption for test data of scenario 2.

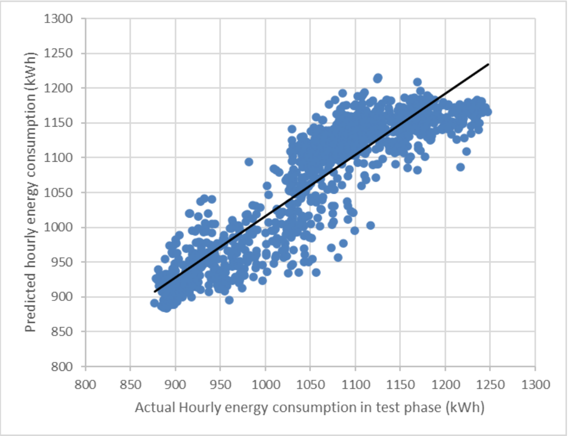

Figure 5 Scatter plot of predicted vs actual hourly energy consumption in the test phase of scenario 2

The graphs generated for the test data set of scenario 2 predicted from 1200 data points using the weather file of Kolkata to generate the scatter plot. Although the scatter plot concentrates along the 45-degree line, the distribution is not uniform, like in scenario 1. The observation is as per expectation because the ML model was tested for a dataset it was not trained with. However, the results are promising since MBE falls within the permissible limits of 10% and CV less than 30% (Yang et al. 2015). This shows that there is scope for further evaluating the use of ML models such as RFR for predicting energy consumption for hourly resolution under changing climatic conditions.

Scenario 1 and Scenario 2 results indicate that the RFR model is a suitable ML model for predicting energy consumption on an hourly basis due to an R^2 value of more than 0.8 in both cases and CV values of less than 10%. The RMSE value is 42.7% in scenario 2, while that in scenario 1 is 28.9%. This indicates a greater spread of the data points in scenario 2 from the mean value Figure 5, corresponding to scenario 2, also indicates the same. The MSE and MBE values for Scenario 2 are also greater than Scenario 1, indicating greater prediction accuracy in Scenario 1.

Conclusions

The study was conducted to demonstrate ML models' capability to predict commercial buildings' energy consumption in hourly resolution for annual time horizons under varying climatic conditions. An archetype reference commercial building model for an 8-hour working schedule representative of the Indian office buildings was used for this study. A methodology was also developed to generate a dataset and predict hourly energy for changing climatic conditions. Four ML models were used: MLR, PR, RFR, and ANN. The model performance evaluation on these four models identified RFR as the most accurate predictive ML model for hourly energy predictions. Thus, the RFR model was used for further evaluation in Scenario 1 and Scenario 2. Scenario 1 was conducted to test the holistic prediction capability of the RFR model and the results of Scenario 1 indicate that the RFR model is suitable for hourly energy prediction. Scenario 2 was conducted to check the capability of the ML model to predict hourly energy consumption for test data for which it was not trained. Model performance evaluation results for scenario 2, although less accurate than scenario 1, are promising. This indicates that there is scope for using ML models such as RFR for predicting hourly energy consumption and changing weather conditions.

The comparison of Scenarios 1 and 2 results indicates that the relationship between input weather features and building energy output needs further study. Further studies in this direction can consider ensemble-based methods, which combine the output of multiple machine-learning algorithms to enhance the prediction performance of a single data-driven model. Commonly used ensemble methods such as bagging, boosting, and stacking may be deployed to improve the hourly energy prediction for changing weather conditions such as the one considered in this experiment in scenario 2. This will further aid in improving the efficiency of building operation systems and support the transition of existing buildings to achieve net-zero energy.

Acknowledgments

The authors acknowledge the funding support from IC-IMPACTS (India-Canada Centre for Innovative Multidisciplinary Partnerships to Accelerate Community Transformation and Sustainability) and DST (Department of Science and Technology), Government of India, grant number DST/IC/IC-IMPACTS/2022/P-7, as well as the Canada Research Chairs Program.

References

Bhatnagar, M., Mathur, J., & Garg, V. (2019). Development of reference building models for India. Journal of Building Engineering, 21, 267–277. https://doi.org/10.1016/j.jobe.2018.10.027

Bhatnagar, M., Mathur, J., Garg, V., & Iqbal, J. (2018). Development Of a Method for Selection of Representative City in A Climate Zone. www.ashrae.org

Chan, J., Shen, J., Valdes, O., Drope, L., & Frisque, A. (2022). An Automated Approach to Parametric Design Analysis for Large-scale Design Projects. Building Simulation Conference Proceedings, 3372–3378. https://doi.org/10.26868/25222708.2021.31080

Ciulla, G., & D'Amico, A. (2019). Building energy performance forecasting: A multiple linear regression approach. Applied Energy, 253. https://doi.org/10.1016/j.apenergy.2019.113500

Wang, L., & Liu, X. (2017). Prediction of the impacts of climatic change on energy consumption for a medium-sized office building with two climate models. Energy and Buildings, 157

Eco-Niwas Samhita. (2018). Energy Conservation Building Code for Residential Buildings, Bureau of Energy Efficiency (BEE), Government of India

India Energy Outlook (2021) International Energy Agency (IEA). World Energy Outlook Special Report. www.iea.org/t&c/

National Building Code of India, Volume 2 (2016) Bureau of Indian Standards (BIS)

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modeling and energy performance prediction: A review. In Renewable and Sustainable Energy Reviews (Vol. 23,pp.272–288). https://doi.org/10.1016/j.rser.2013.03.004

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. Journal of Parallel and Distributed Computing, 134, 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007

Kumar, S. (2016). Performance Based Rating and Energy Performance Benchmarking for Commercial Buildings In India. https://www.researchgate.net/publication/266225743

Miller, C., & Meggers, F. (2017). The Building Data Genome Project: An open, public data set from non-residential building electrical meters. Energy Procedia, 122, 439–444. https://doi.org/10.1016/j.egypro.2017.07.400

Mosavi, A., Salimi, M., Ardabili, S. F., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. In Energies (Vol. 12, Issue 7). MDPI AG. https://doi.org/10.3390/en12071301

Ngo, N. T. (2019). Early predicting cooling loads for energy-efficient design in office buildings by machine learning. Energy and Buildings, 182, 264–273. https://doi.org/10.1016/j.enbuild.2018.10.004

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. Journal of Building Engineering, 45. https://doi.org/10.1016/j.jobe.2021.103406

Pham, A. D., Ngo, N. T., Ha Truong, T. T., Huynh, N. T., & Truong, N. S. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. Journal of Cleaner Production, 260. https://doi.org/10.1016/j.jclepro.2020.121082

Energy Statistics India. (2022). Ministry of Statistics and Program Implementation (MoSPI), Government of India

Sun, Y., Haghighat, F., & Fung, B. C. M. (2020). A review of the state-of-the-art in data-driven approaches for building energy prediction. In Energy and Buildings (Vol. 221). Elsevier Ltd. https://doi.org/10.1016/j.enbuild.2020.110022

Tiberiu Catalina, V. I. B. C. (2012). Multiple regression model for fast prediction of the heating energy demand.

Yang, Z., & Becerik-Gerber, B. (2015). A model calibration framework for simultaneous multi-level building energy simulation. Applied Energy, 149, 415–431. https://doi.org/10.1016/j.apenergy.2015.03.048

Li, K., Xie, X., Xue, W., Dai, X., Chen, X., & Yang, X. (2018). A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. Energy and Buildings, 174, 323–334. https://doi.org/10.1016/j.enbuild.2018.06.017

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. In Renewable and Sustainable Energy Reviews (Vol. 81, pp. 1192–1205). Elsevier Ltd. https://doi.org/10.1016/j.rser.2017.04.095