Journal of Cleaner Production 260 (2020) 121082

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability

Anh-Duc Pham [a], Ngoc-Tri Ngo [a,*], Thi Thu Ha Truong [b], Nhat-To Huynh [a], Ngoc-Son Truong [a]

[a] Faculty of Project Management, The University of Danang - University of Science and Technology, 54 Nguyen Luong Bang, Danang, Viet Nam
[b] Department of Civil Engineering, The University of Danang – University of Technology and Education, 48 Cao Thang Street, Da Nang City, Viet Nam

ARTICLE INFO

Article history:
Received 6 January 2020
Received in revised form
10 March 2020
Accepted 11 March 2020
Available online 14 March 2020

Handling editor: Cecilia Maria Villas Bôas de Almeida

Keywords:
Building energy consumption
Machine learning
Energy efficiency
Random forests
Short-term forecast

ABSTRACT

Buildings must be energy efficient and sustainable because buildings have contributed significantly to world energy consumption and greenhouse gas emission. Predicting energy consumption patterns in buildings is beneficial to utility companies, users, and facility managers because it can help to improve energy efficiency. This work proposed a Random Forests (RF) – based prediction model to predict the short-term energy consumption in the hourly resolution in multiple buildings. Five one-year datasets of hourly building energy consumption were used to examine the effectiveness of the RF model throughout the training and test phases. The evaluation results presented that the RF model exhibited a good prediction accuracy in the prediction. In four evaluation scenarios, the mean absolute error (MAE) values ranged from 0.430 to 0.501 kWh for the 1-step-ahead prediction, from 0.612 to 0.940 kWh for the 12-steps-ahead prediction, and from 0.626 to 0.868 kWh for the 24-steps-ahead prediction. The RF model was superior to the M5P and Random Tree (RT) models. The RF was better about 49.21%, 46.93% in the MAE and mean absolute percentage error (MAPE) than the RT model in forecasting 1-step-ahead building energy consumption. The RF model approved the outstanding performance with the improvement of 49.95% and 29.29% in MAE compared to the M5P model in the 12-steps-ahead, and 24-steps-ahead energy use, respectively. Thus, the proposed RF model was an effective prediction model among the investigated machine learning (ML) models. This study contributes to (i) the state of the knowledge by examining the generalization and effectiveness of ML models in predicting building energy consumption patterns; and (ii) the state of practice by proposing an effective tool to help the building owners and facility managers in understanding building energy performance for enhancing the energy efficiency in buildings.

1. Introduction

Energy demand has significantly increased in the building sector over the past decades because of the increased population, rapid urbanization, and social demand (Amasyali and El-Gohary, 2018). Buildings have contributed significantly to world energy consumption and greenhouse gas emission (Allouhi et al., 2015). Thus, buildings must be energy efficient and sustainable. Understanding energy consumption patterns in buildings is beneficial to utility companies, users, and facility managers because it can help to improve energy efficiency. Research on energy-saving and building energy efficiency has attracted the attention of researchers and practitioners (Guo et al., 2018; Foucquier et al., 2013; Killian and Kozek, 2016). With the development of industry 4.0, insights behind data are enabled to investigate and explore using artificial intelligence (AI) - based methods such as machine learning (ML) techniques. Therefore, predicting the energy consumption patterns in buildings is vital to improving building energy performance during the operation and maintenance stage.

Building energy patterns can be analyzed using the physics-based simulation approach or the data-driven approach (Amasyali and El-Gohary, 2018). The whole building energy simulation tools such as DOE-2 and EnergyPlus are often used for the former approach. These tools require users to provide a large

* Corresponding author.
E-mail addresses: paduc@dut.udn.vn (A.-D. Pham), trinn@dut.udn.vn (N.-T. Ngo), tttha@ute.udn.vn (T.T. Ha Truong), hnto@dut.udn.vn (N.-T. Huynh), tnson@dut.udn.vn (N.-S. Truong).

https://doi.org/10.1016/j.jclepro.2020.121082
0959-6526/

amount of detail information including material thermal properties of building elements such as walls, roofs, window, an HVAC (heating, ventilation, and air conditioning) system, a building geometry, thermal settings, internal loads of occupancy, lighting, equipment, and weather conditions. However, these tools are limited for energy analysis during the operation and maintenance stages due to a lack of the required information and time-consuming (Amasyali and El-Gohary, 2018). Notably, the complexity and difficulty of using this approach are increasing when performing energy analysis for multiple buildings such as at the city scale.

For the latter approach, building energy performance is accessed mainly based on historical data. With the emergence of the AI and the Internet of Things (IoT), there is an opportunity to understating building energy consumption patterns using this approach. The availability of open data is imperative to apply this data-driven approach. Various studies have proposed ML-based prediction models to forecast future uses of building energy. For example, the ML models applied in building energy prediction mainly include the artificial neural networks (ANNs) (Wu et al., 2007), the decision tree (DT) (Killian and Kozek, 2016), adaptive neuro-fuzzy inferring system (ANFIS) (Xie et al., 2019), and the support vector machine (SVM) (Wu et al., 2007).

The efficient use of energy could reduce energy demand, thus increasing monetary savings, reducing greenhouse gas, and improved energy security (McNeil et al., 2019). Ocampo Batlle et al. (2020) developed energy baselines and energy performance indicators to investigate energy reduction potentials in buildings (Ocampo Batlle et al., 2020). Their findings revealed that potential annual energy savings for the educational buildings were up to 9.6%. A method was developed to quantitatively predict the effect of greening on building energy consumption to support decision-makers for implementing urban greening policy (Qiao and Liu, 2020).

Most of the proposed prediction models in previous works have been trained and evaluated using a single dataset or few datasets. This may result in an inadequate generalization of those models in the predictions. To the best of the authors' knowledge, few studies used multiple datasets in evaluating the prediction performance of models in predicting hourly building energy consumption. To fill this gap, this work aims to investigate the generalization and effectiveness of ML models in the prediction of future energy consumption in buildings using multiple datasets from experiments.

Among ML techniques, the Random Forest (RF) is fast in the training process and powerful for solving high-dimensional data and complex problems in the industry (Zhu et al., 2020). Its performance is stable and accurate due to it creates multiple decision trees and combines them to produce output. Particularly, the RF, M5P, random tree (RT) have adopted this study to predict short-term energy consumption in buildings because of their wide use in various fields. The 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy consumption in buildings were predicted in this study. The novelty of this study is to investigate the effectiveness of the RF model for predicting energy consumption in various buildings. In addition, the 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy consumption in buildings were predicted in this study. The investigated ML models were trained and tested using five datasets that collected hourly in one year. Model performance was assessed using statistical measures that consist of mean absolute percentage errors (MAPE), mean absolute errors (MAE), root-mean-square errors (RMSE), and synthesis index (SI).

This study contributes to (i) the state of the knowledge by examining the generalization and effectiveness of ML models in predicting building energy consumption patterns; and (ii) the state of practice by proposing an effective tool to help building owners and facility managers in understanding building energy performance for enhancing the energy efficiency in buildings.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature, and Section 3 presents the methodology. Section 4 describes a dataset and evaluation measures. Section 5 reveals evaluation results and Section 6 discusses the findings. Section 7 provides the concluding remarks and an outline for future work.

2. Literature review

Prediction of energy usage in buildings is becoming increasingly important for energy management, efficiency improvement of lighting, equipment, and HVAC system, cooperation between facility managers and power suppliers (Tian et al., 2019). The statistical models have been applied in building energy use predictions (Kneifel and Webb, 2016). Various ML techniques have been proposed to estimate energy use patterns in buildings such as adaptive learning-based models (Guo et al., 2018) and ANNs models (Chou and Bui, 2014). ML-based prediction models such as the Linear Regression, RF, support vector regression (SVR), and deep neural network were examined in Fan et al. (2017) to forecast cooling loads in buildings.

Short-term load prediction is necessary for motioning and controlling the operation of the HVAC system. A prediction model was introduced for building cooling loads coupled an ANNs with an ensemble approach. The ANNs models have been widely applied for predicting load forecasting problem because of its non-linear characteristics (Singh and Dwivedi, 2018). A prediction model was developed that integrated the ANNs model with an evolutionary optimization algorithm to enhance the prediction accuracy (Singh and Dwivedi, 2018). Their hybrid model was effective in solving short-term load forecast problems because the parameters of the ANN models were fine-tuned during the prediction.

A hybrid nature-inspired optimization algorithm was developed to modeling energy consumption in buildings. Their hybrid model was developed by combining the Autoregressive Integrated Moving Average (ARIMA), SVR, and Particle Swarm Optimization (PSO) (Goudarzi et al., 2019). Goudarzi et al. (2019) proposed a non-intrusive occupant load monitoring (NIOLM) approach to link occupant behaviors and energy consumption in buildings. Particularly, their approach integrated occupancy-sensing data with power changes in energy data to disaggregate building-wide data down to the individual (Rafsanjani et al., 2018).

Because of the limitation of the availability of building energy consumption data, Tian et al. (2019) proposed the parallel prediction scheme using Generative Adversarial Nets (GAN). The GAN was used to generate parallel data from a small amount of the original data. Then, the original data and artificial data were mixed to train the prediction models. Their proposed method was confirmed as the most effective model compared to the methods such as the information diffusion technology and the bootstrap method (Tian et al., 2019).

Among ML models, RF models have been known as a potential and powerful ML technique in solving regression problems (Wang et al., 2018). Notably, its effectiveness was examined in predicting hourly energy consumption in educational buildings (Wang et al., 2018). Comparison results revealed that the predictive accuracy of the RF model improved 14–25% and 5–5.5% compared to the regression tree and SVR models. A literature review indicates that ML models have been used for forecasting building energy, and few studies have presented the effectiveness of ML models in predicting energy consumptions in multiple buildings. Meanwhile, RF models are powerful in solving regression problems Thus, this study attempts to examine the RF model in predicting the short-term energy consumption in buildings.

3. Methodology

3.1. Random forests

The RF was proposed by Breiman (2001), which is an ensemble learning method. The RF model is recognized as a powerful ML (Qiu et al., 2017). The RF integrates multiple decision trees to mitigate the variance of the model without increasing the bias. The RF combines bagging and a random subset of features. Oblique RF models via a least-square estimation were proposed in Qiu et al. (2017) to forecast time series data.

The RF model is an ensemble of decision trees method, that is fast and robust to the noise of data (Kontschieder et al., 2011). The combination of trees in the RF can reduce the error in solving regression problems because it uses the bootstrap aggregation or bagging. The main advantage of the RF is to reduce the predictive error by considering all decision trees in the forest and their prediction correlation (Chan and Paelinckx, 2008).

Giving an RF is an ensemble of C trees $T_1(X)$, $T_2(X)$, ..., $T_C(X)$, where $X = x_1, x_2, \ldots, x_m$ is an m-dimension vector of inputs. The resulting ensemble produces C outputs is defined as

$$Y_{pred\_1} = T1(X), Y_{pred\_2} = T_2(X), \ldots, Y_{pred\_C} = T_C\ (X) \tag{1}$$

where Y_{pred_C} is the prediction value obtained by decision tree number C.

The output of all these randomly generated trees is aggregated to obtain a final prediction Y_{pred_C} which is the average values of all trees in the forest. The RF generates C number of decision trees from N number of training samples. For each tree in the forest, bootstrap sampling is performed to create new training sets while the unselected samples are called out-of-bag sets (Jiang et al., 2009). The bootstrap sampling is a resampling method by independently sampling with replacement from original data with the same sample size. The new training set is then used to fully grow an unpruned regression (or classification) tree. By using the random feature selection, a small number of m features (input variables) are randomly selected in each split of a decision tree node. The process is repeated until M decision trees are grown to form a randomly generated forest.

Fig. 1 presents the structure of the RF model for predicting energy consumption in buildings. The hourly building energy data were divided into the training data and the test data. The effectiveness of the RF model in building energy prediction was examined through the training and test processes. The training process of a randomly generated forest are described as below:

1. Resampling data using the bootstrap method from the training dataset:
2. Grow a tree for each above bootstrap sample in which the best split among a randomly selected subset of input variables (number of features - *mtry*), which is the tuning parameter of the RF.
3. Repeat steps 1 and 2 until C trees were grown.

3.2. M5 Model trees — M5P

An M5 model tree, developed by Quinlan (1992) (Quinlan, 1992), is a binary decision tree which has linear regression functions at the terminal nodes. The node is known as the leaf which sets a relationship between independent and dependent variables. Building a tree consists of two steps (Quinlan, 1992; Abdelkader et al., 2015). Fig. 2 demonstrates the schema of an M5 model tree that split the input space into four linear regression functions at the nodes (LM1 through LM4).

1. Dividing data into subsets to build the decision tree. The splitting criterion for the M5 model tree is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node and calculating the expected reduction in this error as a result of testing each attribute at that node. The standard deviation reduction (SDR) is presented in Eq. (2) as follows:

$$RSD = sd(X) - \sum \frac{|X_i|}{|X|} sd(X_i) \tag{2}$$

where X denotes a set of examples that reaches the node, X_i denotes the subset of examples that have the i-th value of the potential set, and sd is the standard deviation.

2. Pruning and smoothing the tree. The overgrown tree is pruned to avoid overfitting and pruned subtrees are replaced by using linear regression functions.

3.3. Random tree

A RT is a tree that is formed by a stochastic process. The RT models can work with both classification and regression problems. The RF operator is similar to classification and regression tree (CART) but it selects a random subset of attributes before it is applied. The size of the subset is defined by the parameter of the subset ratio.

In the RT model, one input variable corresponds to each interior node of the tree. The number of edges of an interior node equals the number of possible values of the corresponding input variable. Each node (or leaf) denotes a value of the label that is given by the values of the input attributes via the path from the root to the leaf. After the RT model is constructed, the pruning technique is carried out to convert an over-fitted tree to a more general form so that the RT's predictive power on unseen datasets is enhanced.

4. Datasets and evaluation process

4.1. Datasets

The effectiveness of the proposed ML-based prediction model was evaluated using published open data. The five datasets from five buildings in the hourly resolution were derived from the building data genome project (Miller and Meggers, 2017). Table 1 presents information on these five buildings including the descriptive statistics of hourly energy consumption data. Fig. 3 illustrates the energy consumption profiles of these five buildings for one year in the hourly resolution. The profiles exposure highly random patterns.

4.2. Evaluation settings

In this study, the RF model was proposed to predict the 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy consumption in buildings in the hourly resolution. The effectiveness and generalization of the investigated RF model were evaluated using five different datasets from five buildings. The evaluation process was performed by four scenarios which used a 3-month dataset, a 6-month dataset, a 9-month, and 12-month datasets.

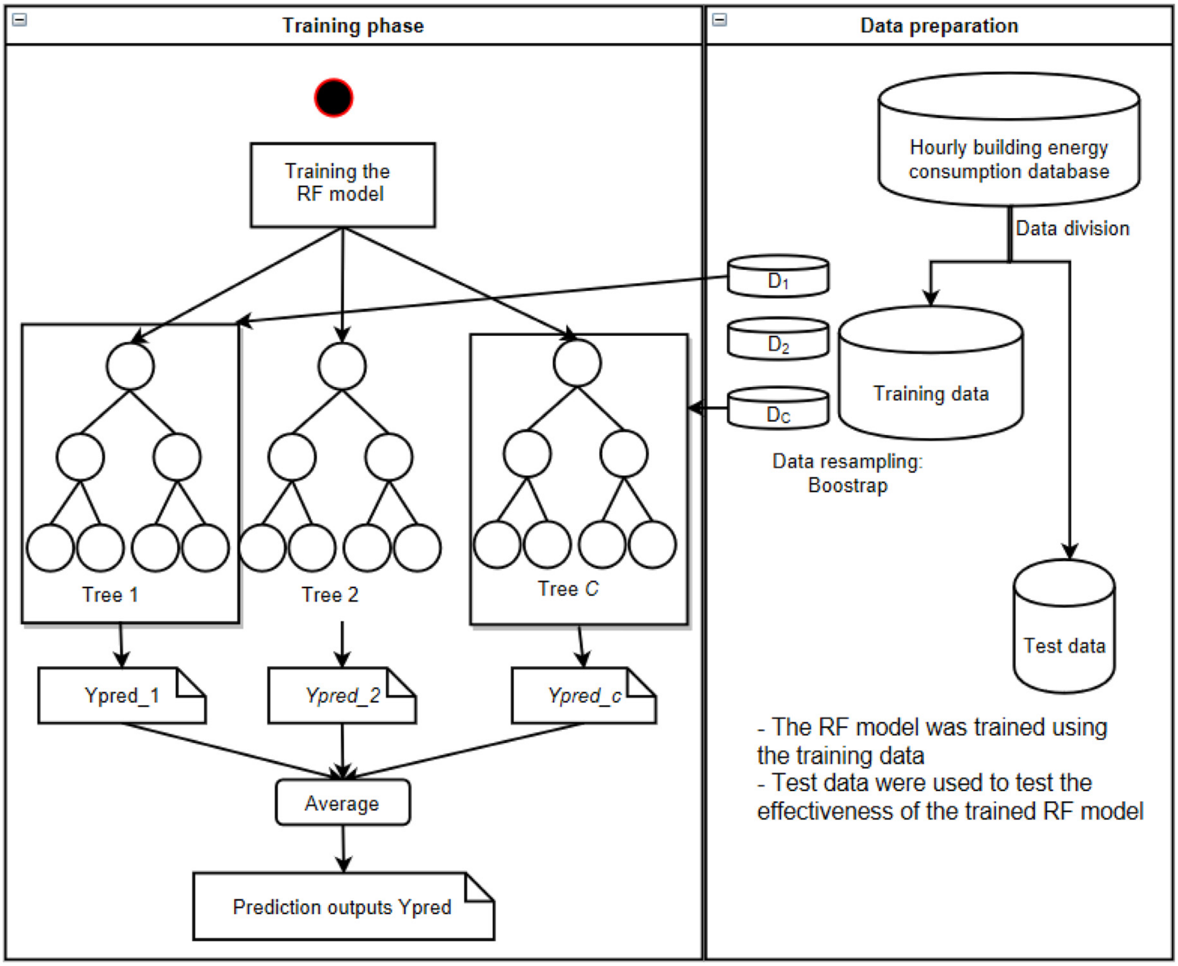


Fig. 1. The random forests model for hourly building energy prediction.

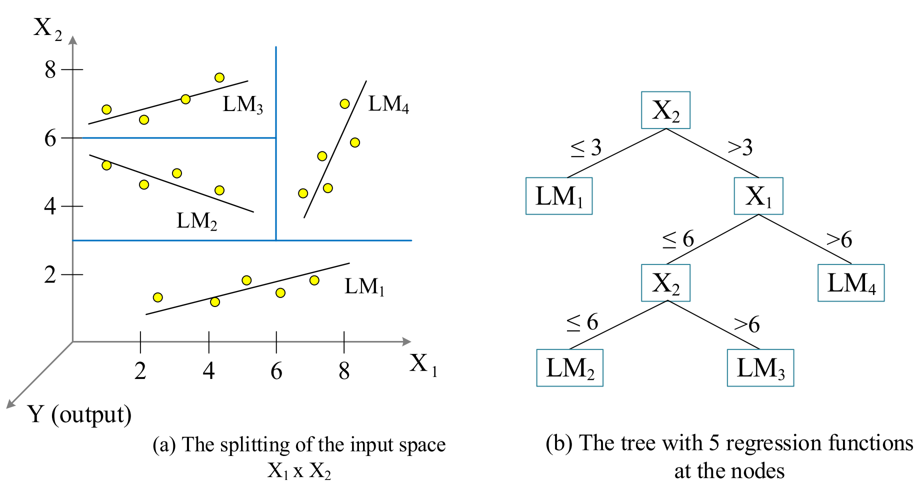


Fig. 2. The schema of a typical M5 tree.

Table 2 presents the sizes of datasets used for training and testing the RF model. For example, in scenario 1, a 3-month hourly energy consumption dataset was used to evaluate the predictive performance of the RF model in which 2-month data (1440 data points) were used to train the RF model and 1-month (720 data points) were used to test the model performance. Fig. 4 visualizes these scenarios in which the training data were used to train or build the ML models while the test data were used to test the effectiveness and generalizability of the trained ML models.

The MAE, the RMSE, the MAPE are common statistical measures used in this study to assess the predictive performance of the RF model. Besides, the SI is used as a normalization measure which is

Table 1
Building energy consumption datasets.

Dataset	Building	Duration	Hourly energy consumption in buildings (kWh)			
			Mean	Minimum	Maximum	Std. dev.
Dataset 1	Office_Cristina	12 months	3.332	1.000	6.475	1.067
Dataset 2	Office_Conrad	12 months	8.973	4.475	16.125	2.307
Dataset 3	Office_Pamela	12 months	27.842	10.037	58.400	11.273
Dataset 4	Office_Allyson	12 months	19.717	8.770	65.623	9.768
Dataset 5	Office_Amelie	12 months	28.711	1.570	98.898	22.199

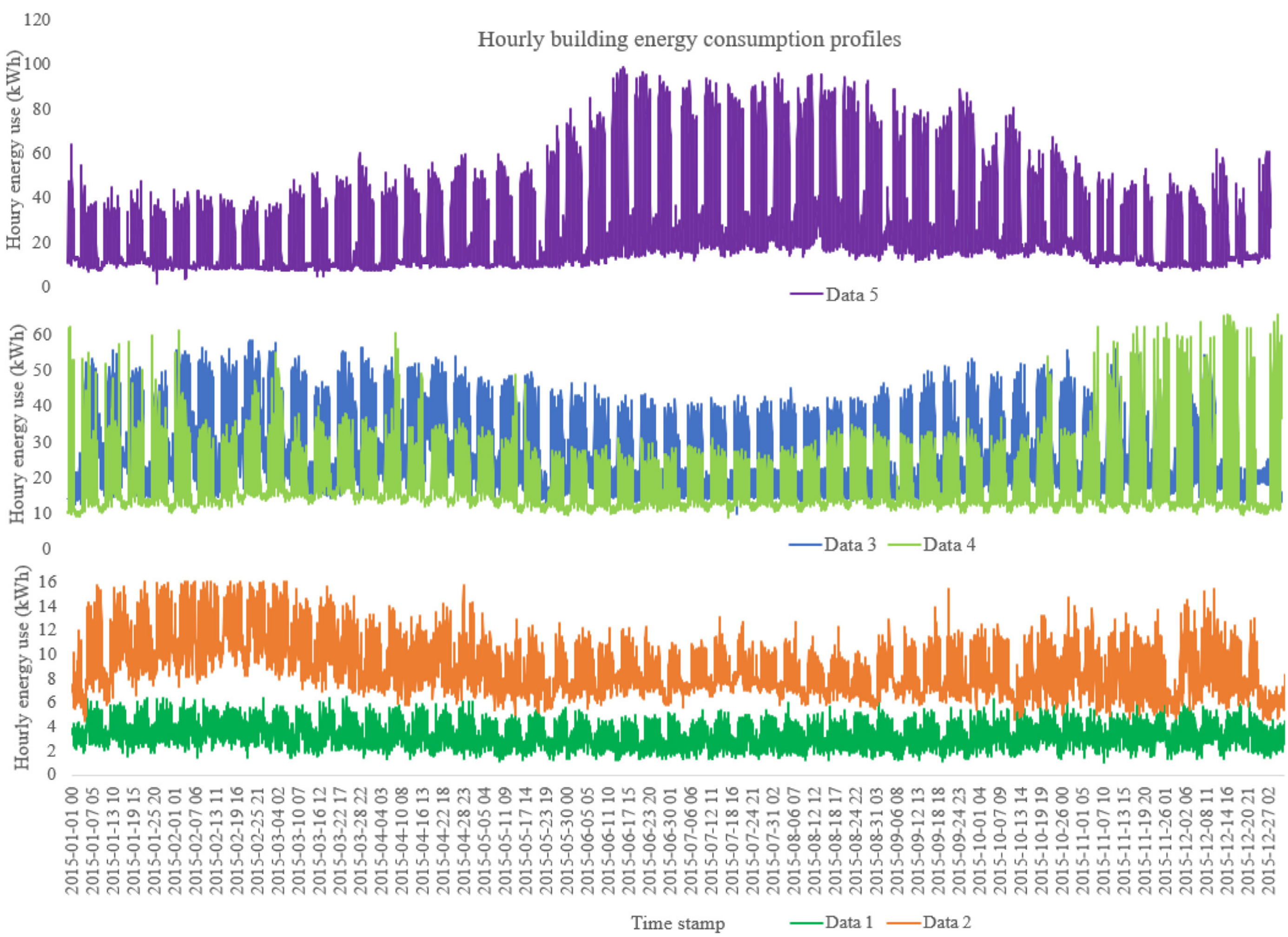


Fig. 3. Historical hourly energy consumption profiles in buildings for the one year.

Table 2
Scenario of data division for evaluating RF models.

Scenario	Data size	Training data	Test data
1	3-month dataset	2-month dataset ~1440 data	1-month dataset ~720 data
2	6-month dataset	5-month dataset ~3600 data	1-month dataset ~720 data
3	9-month dataset	8-month dataset ~5760 data	1-month dataset ~720 data
4	12-month dataset	11-month dataset ~7920 data	1-month dataset ~720 data

calculated based on the MAE, RMSE, and MAPE values. Equations (3) – (6) present the mathematical form of these statistical measures, which are determined based on the actual data and predicted data. The evaluation process was performed using the open-source WEKA data mining tool (Hall et al., 2009). The parameters of the RF model were set as defaults (Table 3). The batch size was set as 100 that is recommended by the data mining tool (Hall et al., 2009). It means the total dataset in the training data was used to train the prediction model.

Table 3
Default settings of parameters of the Random Forests model.

Parameters	Description	Setting
bagSizePercent	Size of each bag, as a percentage of the training set size.	100
batchSize	The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.	100
numFeatures	Sets the number of randomly chosen attributes. If 0, int(log_2(#predictors) + 1) is used.	0
numIterations	The number of iterations to be performed.	100
Seed	The random number seed to be used.	1
maxDepth	The maximum depth of the tree, 0 for unlimited.	0

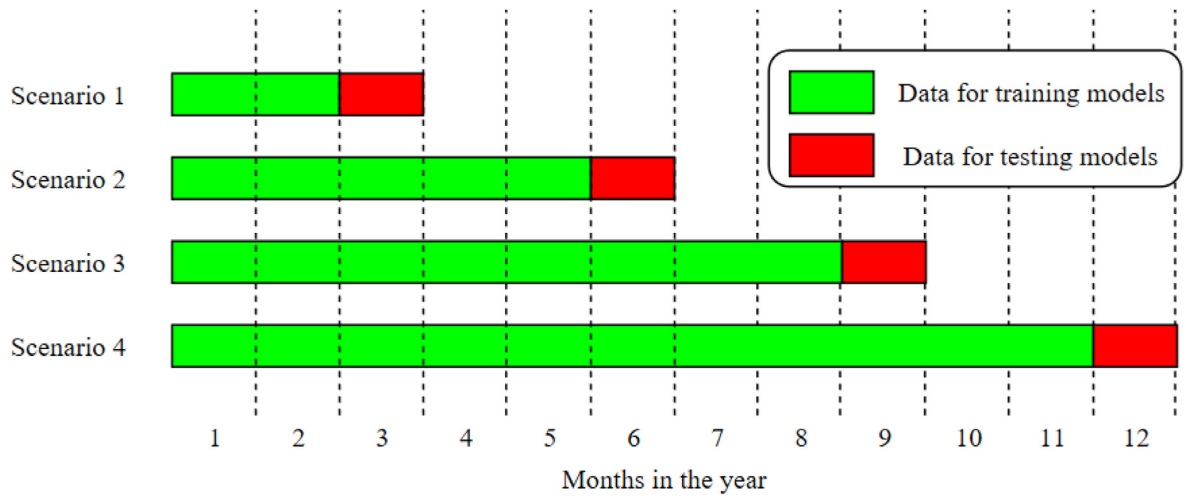


Fig. 4. Data division for training and testing prediction models.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y - y'}{y}\right| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y')^2} \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y - y'| \quad (5)$$

$$SI = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}}\right) \quad (6)$$

where y' and y are = predicted and actual hourly energy consumption data, respectively; n = size of the data; m = number of performance measures; and P_i = ith performance measure. $P_{\min,i}$is the minimum values ith performance measure, $P_{\max,i}$is the maximum values ith performance measure.

5. Analytical results

The effectiveness of the RF model was evaluated in predicting 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy consumptions in buildings. The predictive performance of the RF models was examined using three statistical measures including MAE, MAPE, and RMSE. The RF model was trained and tested using five different datasets. Table 4 summarizes the predictive performance of the RF model in predicting 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy consumptions in buildings that obtained from the dataset 1 regarding four evaluation scenarios during the training and test phases.

In the training phase of four evaluation scenarios of building energy use predictions, as shown in Table 4, the MAE values obtained by the RF model were ranged from 0.158 to 0.166 kWh for the 1-step-ahead prediction, from 0.491 to 0.502 kWh for the 12-steps-ahead prediction, and from 0.623 to 0.689 kWh for the 24-steps-ahead prediction. As shown in Table 4, at the test phase, the RF model exhibited a good prediction accuracy in the prediction. In four evaluation scenarios, the MAE values ranged from 0.430 to 0.501 kWh for the 1-step-ahead prediction, from 0.612 to 0.940 kWh for the 12-steps-ahead prediction, and from 0.626 to 0.868 kWh for the 24-steps-ahead prediction. Fig. 5 visualized the predicted and actual values of the 1-step-ahead energy consumption predictions for dataset 1 in the test phase. The black line indicates a perfect agreement between the actual and predicted energy data. Fig. 5 reveals a good agreement between predicted and actual data that confirmed the effectiveness of the RF model in forecasting time-series energy usage in buildings.

To provide readers with a comprehensive analysis, this study explored the performance variation of the RF model among four evaluation scenarios obtained by using five datasets. This analysis aims to identify under which scenario or the length of the training data that brings a competitive accuracy in the prediction. The lengths of the training data were different among the four scenarios as shown in Table 1. The SI values were used as a comparing index among the four scenarios, which was calculated based on the MAE, MAPE, and RMSE values.

Tables 5–9 presents the predictive performance of the RF model in predicting in 1-step-ahead energy use, 12-steps-ahead energy use, 24-steps-ahead energy use regarding five investigated test data in the test phase. For dataset 1, the comparison results reveal

Table 4
Prediction accuracy of the RF models for dataset 1 during the training and test phases.

RF-based prediction model for dataset 1		Training phase			Test phase		
Scenario	Performance measure	1-step-ahead	12-steps-ahead	24-steps-ahead	1-step-ahead	12-steps-ahead	24-steps-ahead
Scenario 1	MAE (kWh)	0.166	0.491	0.689	0.501	0.940	0.868
	MAPE (%)	4.813	13.198	18.196	16.832	27.670	25.548
	RMSE (kWh)	0.210	0.712	0.942	0.615	1.127	1.056
Scenario 2	MAE (kWh)	0.163	0.501	0.647	0.436	0.806	0.772
	MAPE (%)	5.189	15.189	19.621	17.108	27.533	26.413
	RMSE (kWh)	0.209	0.721	0.889	0.549	0.982	0.958
Scenario 3	MAE (kWh)	0.158	0.498	0.626	0.432	0.663	0.652
	MAPE (%)	5.375	16.171	20.262	16.804	25.293	25.258
	RMSE (kWh)	0.203	0.716	0.871	0.546	0.860	0.861
Scenario 4	MAE (kWh)	0.160	0.502	0.623	0.430	0.612	0.626
	MAPE (%)	5.451	16.866	20.978	14.255	20.752	21.155
	RMSE (kWh)	0.204	0.717	0.868	0.547	0.814	0.827

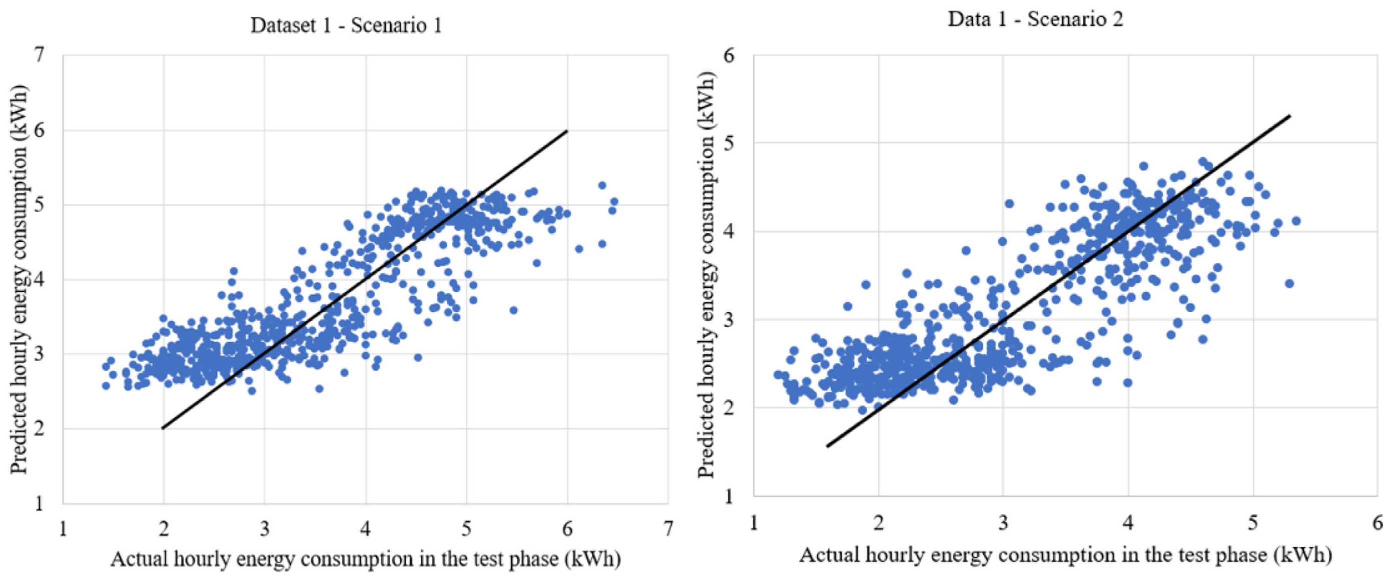


Fig. 5. Scatter of actual and predicted values in the 1-step-ahead predictions in the test phase using the dataset 1.

Table 5
Performance of the RF model obtained by dataset 1 in the test phase.

Dataset	Predictions	Performance	Scenario				Range	
			S1	S2	S3	S4	Min	Max
Dataset 1	1-step-ahead	MAE (kWh)	0.501	0.436	0.432	0.430	0.430	0.501
		MAPE (%)	16.832	17.108	16.804	14.255	14.255	17.108
		RMSE (kWh)	0.615	0.549	0.546	0.547	0.546	0.615
		SI	0.968	0.375	0.304	**0.003**		
		Ranking	4	3	2	**1**		
	12-steps-ahead	MAE (kWh)	0.940	0.806	0.663	0.612	0.612	0.940
		MAPE (%)	27.670	27.533	25.293	20.752	20.752	27.670
		RMSE (kWh)	1.127	0.982	0.860	0.814	0.814	1.127
		SI	1.000	0.702	0.320	**0.000**		
		Ranking	4	3	2	**1**		
	24-steps-ahead	MAE (kWh)	0.868	0.772	0.652	0.626	0.626	0.868
		MAPE (%)	25.548	26.413	25.258	21.155	21.155	26.413
		RMSE (kWh)	1.056	0.958	0.861	0.827	0.827	1.056
		SI	0.945	0.725	0.345	**0.000**		
		Ranking	4	3	2	**1**		

that the RF model obtained the best predictive accuracy at the scenario 4 with the 11-month training data. Under the scenario 4 (11-month training data), the MAPE values were 14.255%, 20.752%, and 21.155% for the 1-step-ahead prediction, the 12-steps-ahead prediction, and the 24-steps-ahead prediction, respectively. These numbers were smaller than those of three remaining scenarios. An observation from the ranking in Table 4 revealed that the performance of the RF model was improved as the increasing length of

Table 6
Performance of the RF model obtained by dataset 2 in the test phase.

Dataset	Predictions	Performance	Scenario				Range	
			S1	S2	S3	S4	Min	Max
Dataset 2	1-step-ahead	MAE (kWh)	1.037	0.499	0.587	0.720	0.499	1.037
		MAPE (%)	10.374	6.023	6.851	9.021	6.023	10.374
		RMSE (kWh)	1.262	0.648	0.842	0.972	0.648	1.262
		SI	1.000	**0.000**	0.223	0.543		
		Ranking	4	**1**	2	3		
	12-steps-ahead	MAE (kWh)	1.633	1.075	1.030	1.213	1.030	1.633
		MAPE (%)	16.710	12.157	12.054	14.571	12.054	16.710
		RMSE (kWh)	1.953	1.477	1.461	1.744	1.461	1.953
		SI	1.000	0.043	**0.000**	0.473		
		Ranking	4	2	**1**	3		
	24-steps-ahead	MAE (kWh)	1.895	0.996	1.079	1.274	0.996	1.895
		MAPE (%)	20.276	11.306	12.556	15.376	11.306	20.276
		RMSE (kWh)	2.320	1.375	1.521	1.841	1.375	2.320
		SI	1.000	**0.000**	0.129	0.419		
		Ranking	4	**1**	2	3		

Table 7
Performance of the RF model obtained by dataset 3 in the test phase.

Dataset	Predictions	Performance	Scenario				Range	
			S1	S2	S3	S4	Min	Max
Dataset 3	1-step-ahead	MAE (kWh)	2.460	2.017	2.155	2.292	2.02	2.46
		MAPE (%)	8.953	8.685	8.476	9.316	8.48	9.32
		RMSE (kWh)	3.266	2.780	3.170	3.081	2.78	3.27
		SI	0.856	**0.083**	0.371	0.746		
		Ranking	4	**1**	2	3		
	12-steps-ahead	MAE (kWh)	6.243	5.671	5.282	4.537	4.54	6.24
		MAPE (%)	22.523	19.840	19.335	17.850	17.85	22.52
		RMSE (kWh)	8.052	7.527	7.879	6.408	6.41	8.05
		SI	1.000	0.590	0.550	**0.000**		
		Ranking	4	3	2	**1**		
	24-steps-ahead	MAE (kWh)	6.231	5.982	5.367	5.328	5.33	6.23
		MAPE (%)	22.793	22.235	20.046	21.614	20.05	22.79
		RMSE (kWh)	7.986	7.401	7.821	7.204	7.20	7.99
		SI	1.000	0.591	0.277	**0.190**		
		Ranking	4	3	2	**1**		

Table 8
Performance of the RF model obtained by dataset 4 in the test phase.

Dataset	Predictions	Performance	Scenario				Range	
			S1	S2	S3	S4	Min	Max
Dataset 4	1-step-ahead	MAE (kWh)	1.464	1.231	1.112	4.031	1.112	4.031
		MAPE (%)	6.467	7.270	6.909	16.868	6.467	16.868
		RMSE (kWh)	2.278	1.915	1.500	6.096	1.500	6.096
		SI	0.097	0.069	**0.014**	1.000		
		Ranking	3	2	**1**	4		
	12-steps-ahead	MAE (kWh)	4.109	3.583	4.092	13.229	3.583	13.229
		MAPE (%)	21.298	24.753	28.659	90.989	21.298	90.989
		RMSE (kWh)	7.021	6.196	6.536	17.120	6.196	17.120
		SI	0.043	**0.017**	0.063	1.000		
		Ranking	2	**1**	3	4		
	24-steps-ahead	MAE (kWh)	3.957	4.322	3.228	8.721	3.228	8.721
		MAPE (%)	19.341	30.370	20.586	44.433	19.341	44.433
		RMSE (kWh)	6.876	6.547	5.459	13.195	5.459	13.195
		SI	0.105	0.260	**0.017**	1.000		
		Ranking	2	3	**1**	4		

the training data.

The comparison results in Table 6 show that the RF model yielded the best predictive accuracy for predicting 1-step-ahead prediction and 24-steps-ahead predictions in scenario 2 (the training data of 5 months). Particularly, the evaluation results in Table 6 shows that the MAPE values were 6.023% and 11.306% for predicting 1-step-ahead and 24-steps-ahead energy consumption in buildings, respectively. For predicting 12-steps-ahead energy data, scenario 3 brought the best performance with the MAPE of 12.054%, the MAE of 1.030 kWh, and the RMSE of 1.461 kWh. Besides, the performance of the RF model in scenario 2 was quite like that of the RF model in scenario 3.

The performance of prediction models is affected partially by the data size that is used to train models. As a deep observation

Table 9
Performance of the RF model obtained by dataset 5 in the test phase.

Dataset	Predictions	Performance	Scenario				Range	
			S1	S2	S3	S4	Min	Max
Dataset 5	1-step-ahead	MAE (kWh)	3.610	10.176	3.308	3.619	3.308	10.176
		MAPE (%)	21.398	21.175	11.003	21.511	11.003	21.511
		RMSE (kWh)	5.455	15.726	5.526	5.623	5.455	15.726
		SI	0.344	0.989	**0.002**	0.354		
		Ranking	2	4	**1**	3		
	12-steps-ahead	MAE (kWh)	8.319	16.854	11.399	7.516	7.516	16.854
		MAPE (%)	64.097	41.741	29.093	53.354	29.093	64.097
		RMSE (kWh)	11.619	24.406	18.015	11.449	11.449	24.406
		SI	0.366	0.787	0.308	**0.231**		
		Ranking	3	4	2	**1**		
	24-steps-ahead	MAE (kWh)	6.832	16.734	12.868	7.873	6.832	16.734
		MAPE (%)	42.969	35.636	35.802	52.650	35.636	52.650
		RMSE (kWh)	11.075	25.334	18.549	10.805	10.805	25.334
		SI	**0.150**	0.667	0.384	0.368		
		Ranking	**1**	4	3	2		

from the evaluation results showed in Tables 5–9, the proposed RF model performed well the predictions at the scenarios 2, 3, and 4 in which the training data were the 5-month data, the 8-month data, and the 11-month data in the hourly resolution. A scenario with the best accuracy for each dataset was bolded in Tables 5–9. The significant finding from this analysis is that the optimal sizes of the training data were not consistent among five datasets during the predictions. This finding made sense because the patterns of these datasets were differentiated in nature. This result suggests that identifying an optimal size of the training data was necessary to improve the model's performance in predicting energy consumption in a specific building.

Fig. 6 plots the MAPE and MAE values obtained by the RF model among five datasets in predicting the 1-step-ahead prediction (Fig. 6a and b), the 12-steps-ahead prediction (Fig. 6c and d) and the 24-steps-ahead prediction (Fig. 6e and f). The analytical results confirmed that predictive accuracy was different among datasets. Notably, Fig. 6 revealed that the performance was pretty good at the first four datasets.

6. Discussion

To provide readers with a comprehensive analysis, the predictive performance of the RF model was compared to those of some common ML models including M5P and RT models. These models were selected because they were belonging to the tree family-based ML techniques. Table 10 presents the default settings of these models. The comparisons were performed regarding the 1-step-ahead prediction, 12-steps-ahead prediction, and 24-steps-ahead prediction throughout five datasets. Sizes of the training data used for training these compared models were selected base on the best scenario for each dataset as mentioned in Section 5. The predictive performance of these models was ranked based on the SI values that were calculated from MAE, MAPE, and RMSE. The smaller the SI value is, the better the model is.

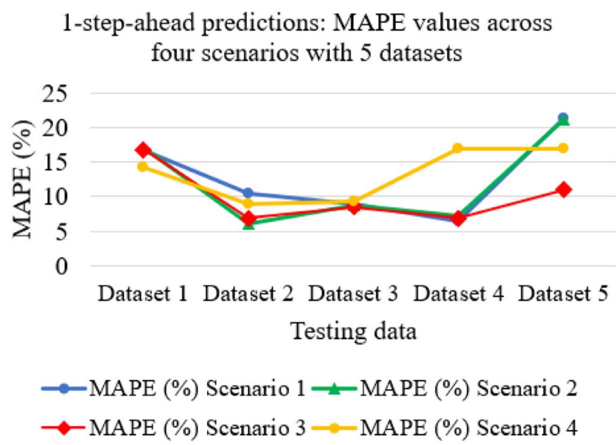
Table 11 presents the performance comparison among the RF, M5P, and RT models in predicting the 1-step-ahead prediction, 12-steps-ahead prediction, and 24-steps- ahead prediction. For predicting the 1-step-ahead building energy consumption, the MAE values obtained by the RF, M5P, and RT models were 0.430 kWh, 0.456 kWh, 0.649 kWh, respectively. The comparative results in Table 11 show that the RF model outperformed the M5P and RT models in terms of all measures (i.e., MAE, MAPE, RMSE, SI) in predicting the 1-step-ahead energy use in buildings. For predicting 12-steps-ahead and 24-steps-ahead energy consumption, the results in Table 11 also revealed that the RF model was superior to the M5P and RT models in terms of the SI values.

Tables 12 – 15 summary performance measures (*i.e.,* MAE, MAPE, RMSE, and SI) obtained by the RF, M5P, and RT prediction models using datasets 2, 3, 4, and 5 during the test phase. The comparison results in these Tables reveal that The RF model yielded the best accuracy comparing to the M5P and RT models in predicting hourly building energy consumption. It was followed by the M5P model with close accuracy. The performance of the RF model was superior significantly to that of the RT model for building energy consumption prediction. For example, for dataset 2, the RF model obtained the smallest MAPE values of 6.023% in the 1-step-ahead prediction of building energy use, followed by the M5P model with the MAPE of 6.331%, and the RT model with the MAPE of 11.268%. The results also revealed that the RT model was not effective in the prediction compared to the RF and M5P models.

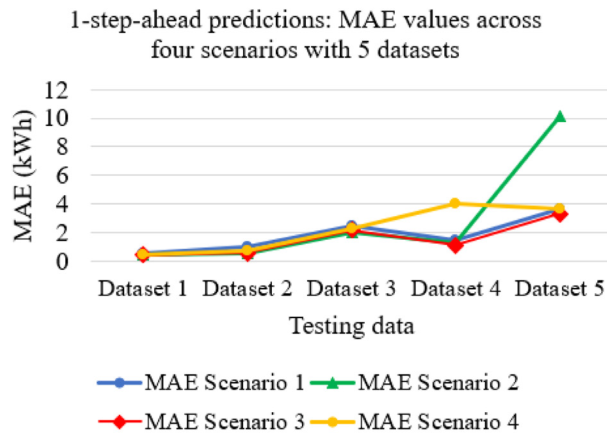
Table 16 depicts the average performance of the comparing models through datasets and performance improvement percentage of the RF model over the M5P and RT models. The ranking results in Table 16 revealed that the RF model was the most effective prediction model among the comparing models with the SI values of 0.014, 0.000, and 0.000 in predicting the 1-step-ahead, 12-steps-ahead, and 24-steps-ahead energy use, respectively. Performance of the RF model in forecasting 1-step-ahead building energy consumption was better about 49.21%, 46.93%, and 69.15% in MAE, MAPE, and RMSE than that of the RT model. The comparison results in Table 14 also depicted that the RF model approved the outstanding performance with the improvement of 49.95% and 29.29% in MAE compared to the M5P model in the 12-steps-ahead, and 24-steps-ahead energy use, respectively. Fig. 7 plots the actual data and predicted data obtained by the RF, M5P, and RT models during a week in the test phase. To sum up, the proposed RF model was an effective prediction model among the investigated ML models.

7. Conclusions

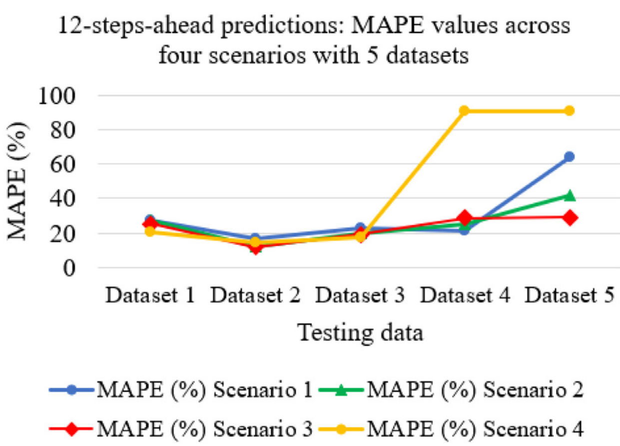
Efficient energy usage in buildings has become a key concern of facility managers for reducing energy consumption during the operation and maintenance. Early forecasting energy use profiles in multiple buildings are vital to providing the building managers with insights to take further actions in reducing energy consumption. This work proposed the random forests – based prediction model in predicting the short-term energy consumption in buildings including the 1-step-ahead, 12-steps-ahead, and 24-steps-



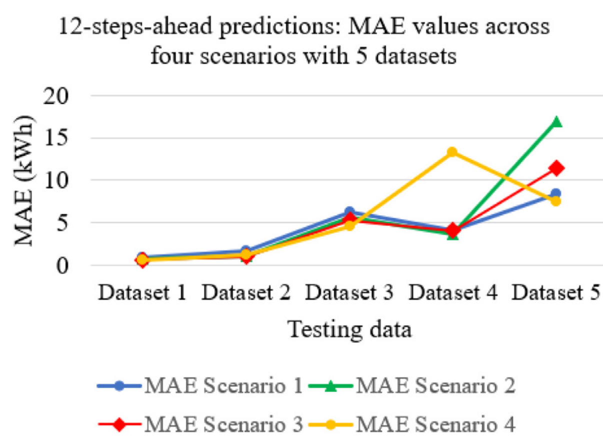
a. MAPE obtianed by 1-step-ahead predictions



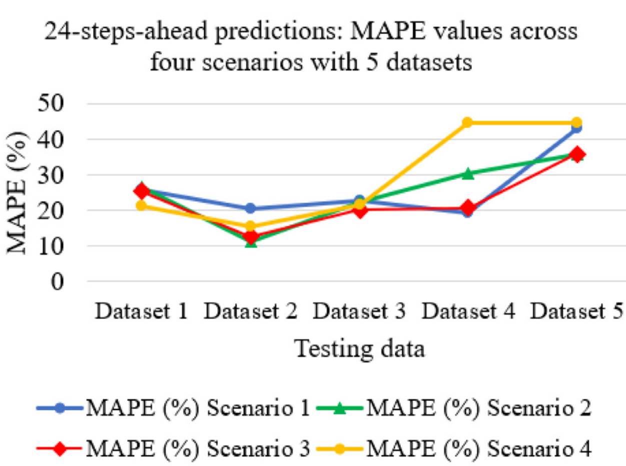
b. MAE obtianed by 1-step-ahead predictions



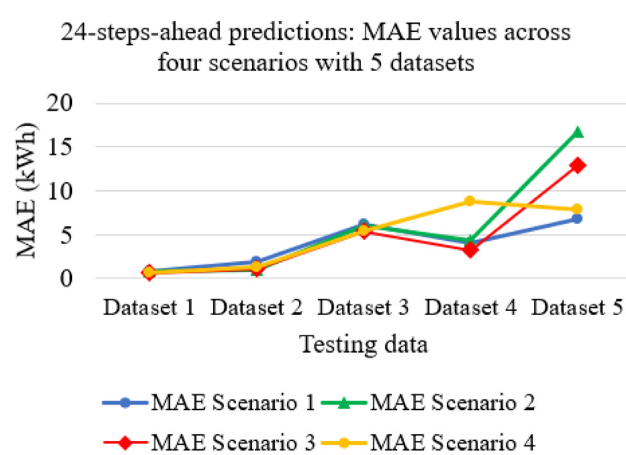
c. MAPE obtianed by 12-step-ahead predictions



d. MAE obtianed by 12-step-ahead predictions



e. MAPE obtianed by 24-step-ahead predictions



f. MAE obtianed by 24-step-ahead predictions

Fig. 6. MAPE and MAE values obtained across the evaluation process.

Table 10
Default setting of the M5P and RT models.

Parameters	Description	Setting M5P	Setting RT
batchSize	The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size.	100	100
buildRegressionTree	Whether to generate a regression tree/rule instead of a model tree/rule.	False	False
minNumInstances	The minimum number of instances to allow at a leaf node.	4.0	4.0
saveInstances	Whether to save instance data at each node in the tree for visualization purposes.	False	False
unpruned	Whether unpruned tree/rules are to be generated.	False	False
useUnsmoothed	Whether to use unsmoothed predictions.	False	False

Table 11
Performance of the compared prediction models for dataset 1 in the test phase.

Dataset	Predictions	Performance measure	Prediction models		
			RF	M5P	RT
Dataset 1	1-step-ahead	MAE (kWh)	**0.430**	0.456	0.649
		MAPE (%)	**14.255**	14.527	20.731
		RMSE (kWh)	**0.547**	0.585	0.817
		SI	**0.000**	0.101	1.000
		Ranking	**1**	2	3
	12-steps-ahead	MAE (kWh)	**0.612**	0.620	0.941
		MAPE (%)	20.752	**20.537**	32.227
		RMSE (kWh)	**0.814**	0.834	1.203
		SI	**0.006**	0.025	1.000
		Ranking	**1**	2	3
	24-steps-ahead	MAE (kWh)	**0.626**	0.652	1.010
		MAPE (%)	**21.155**	21.278	34.055
		RMSE (kWh)	**0.827**	0.863	1.289
		SI	**0.000**	0.052	1.000
		Ranking	**1**	2	3

Table 12
Performance of the compared prediction models for dataset 2 in the test phase.

Dataset	Predictions	Performance measure	Prediction models		
			RF	M5P	RT
Dataset 2	1-step-ahead	MAE (kWh)	0.499	0.526	0.937
		MAPE (%)	**6.023**	6.331	11.268
		RMSE (kWh)	**0.648**	0.687	1.270
		SI	**0.000**	0.061	1.000
		Ranking	**1**	2	3
	12-steps-ahead	MAE (kWh)	**1.030**	1.158	1.460
		MAPE (%)	**12.054**	13.424	17.468
		RMSE (kWh)	**1.461**	1.606	1.888
		SI	**0.000**	0.297	1.000
		Ranking	**1**	2	3
	24-steps-ahead	MAE (kWh)	**0.996**	1.133	1.424
		MAPE (%)	**11.306**	13.342	16.757
		RMSE (kWh)	**1.375**	1.467	1.895
		SI	**0.000**	0.290	1.000
		Ranking	**1**	2	3

Table 13
Performance of the compared prediction models for dataset 3 in the test phase.

Dataset	Predictions	Performance measure	Prediction models		
			RF	M5P	RT
Dataset 3	1-step-ahead	MAE (kWh)	**2.017**	2.086	2.997
		MAPE (%)	8.685	**8.553**	12.205
		RMSE (kWh)	**2.780**	3.027	4.715
		SI	**0.012**	0.066	1.000
		Ranking	**1**	2	3
	12-steps-ahead	MAE (kWh)	**4.537**	4.955	7.196
		MAPE (%)	**17.850**	19.000	28.998
		RMSE (kWh)	**6.408**	7.130	10.127
		SI	**0.000**	0.152	1.000
		Ranking	**1**	2	3
	24-steps-ahead	MAE (kWh)	**5.328**	5.439	7.532
		MAPE (%)	21.614	**21.542**	30.803
		RMSE (kWh)	**7.204**	7.722	10.529
		SI	**0.003**	0.069	1.000
		Ranking	**1**	2	3

Table 14
Performance of the compared prediction models for dataset 4 in the test phase.

Dataset	Predictions	Performance measure	Prediction models		
			RF	M5P	RT
Dataset 4	1-step-ahead	MAE (kWh)	**1.112**	1.120	1.865
		MAPE (%)	6.909	**6.318**	10.748
		RMSE (kWh)	**1.500**	1.583	3.809
		SI	0.044	**0.015**	1.000
		Ranking	2	**1**	3
	12-steps-ahead	MAE (kWh)	**3.583**	11.580	7.103
		MAPE (%)	**24.753**	80.022	46.494
		RMSE (kWh)	**6.196**	23.076	11.952
		SI	**0.000**	1.000	0.391
		Ranking	**1**	3	2
	24-steps-ahead	MAE (kWh)	**3.228**	3.881	5.405
		MAPE (%)	**20.586**	23.022	33.190
		RMSE (kWh)	**5.459**	6.130	8.901
		SI	**0.000**	0.229	1.000
		Ranking	**1**	2	3

Table 15
Performance of the compared prediction models for dataset 5 in the test phase.

Dataset	Predictions	Performance measure	Prediction models		
			RF	M5P	RT
Dataset 5	1-step-ahead	MAE (kWh)	**3.308**	3.166	4.543
		MAPE (%)	11.003	**10.239**	13.920
		RMSE (kWh)	**5.526**	6.090	7.998
		SI	**0.104**	0.076	1.000
		Ranking	**2**	1	3
	12-steps-ahead	MAE (kWh)	**7.516**	7.593	8.971
		MAPE (%)	53.354	**31.135**	52.538
		RMSE (kWh)	**11.449**	12.704	14.825
		SI	**0.333**	0.142	0.988
		Ranking	**2**	1	3
	24-steps-ahead	MAE (kWh)	**6.832**	10.888	13.111
		MAPE (%)	42.969	**40.198**	88.866
		RMSE (kWh)	**11.075**	16.786	18.757
		SI	**0.019**	0.463	1.000
		Ranking	**1**	2	3

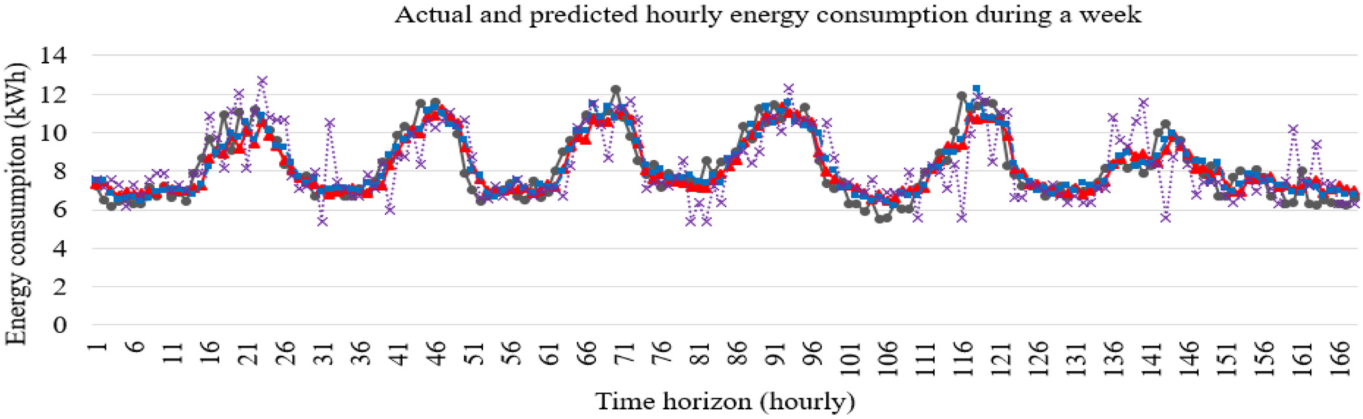
Table 16
Average performance measures of prediction models and performance improvement.

Prediction	Average measure	Prediction model			Performance improvement of RF model (%)	
		RF	M5P	RT	M5P	RT
1-step-ahead prediction	MAE (kWh)	1.473	1.471	2.198	–	49.21
	MAPE (%)	9.375	9.193	13.774	–	46.93
	RMSE (kWh)	2.200	2.395	3.722	8.83	69.15
	SI	**0.014**	0.043	1.000		
	Ranking	**1**	2	3		
12-steps-ahead prediction	MAE (kWh)	3.455	5.181	5.134	49.95	48.58
	MAPE (%)	25.753	32.823	35.545	27.46	38.02
	RMSE (kWh)	5.266	9.070	7.999	72.25	51.91
	SI	**0.000**	0.907	0.897		
	Ranking	**1**	3	2		
24-steps-ahead prediction	MAE (kWh)	3.402	4.398	5.696	29.29	67.45
	MAPE (%)	23.526	23.877	40.734	1.49	73.15
	RMSE (kWh)	5.188	6.593	8.274	27.09	59.49
	SI	**0.000**	0.303	1.000		
	Ranking	**1**	2	3		

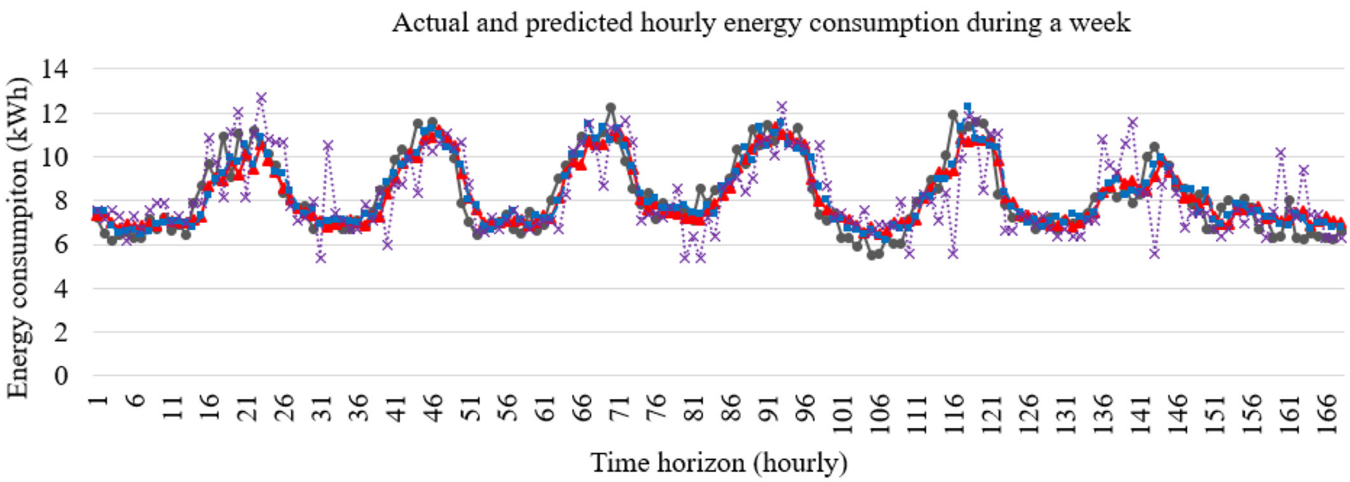
ahead building energy consumption in the hourly resolution.

Five different datasets of the one-year building energy consumption data were used to trained and test the effectiveness of the RF model in predicting the short-term energy uses. Four evaluation scenarios with respect to the length of the learning data were examined the evaluate the predictability of the RF model. Besides, the predictive performance of the RF model was compared to those of M5P and random tree models in predicting the short-term building energy consumption.

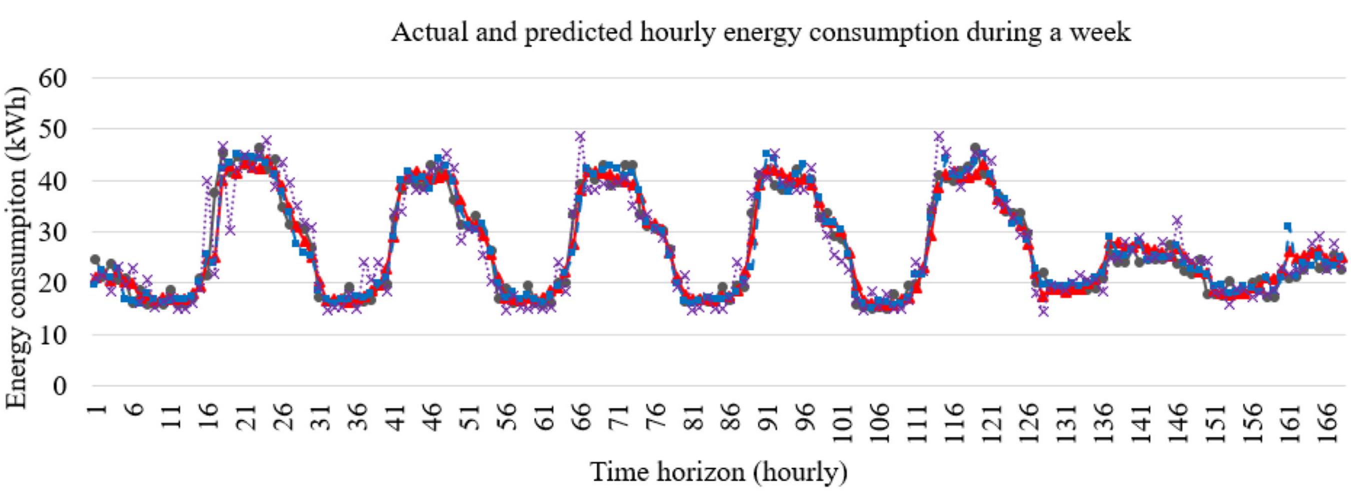
The evaluation results revealed that the RF model was effective in predicting hourly building energy consumption. In four evaluation scenarios, the MAE values ranged from 0.430 to 0.501 kWh for



a. Dataset 1

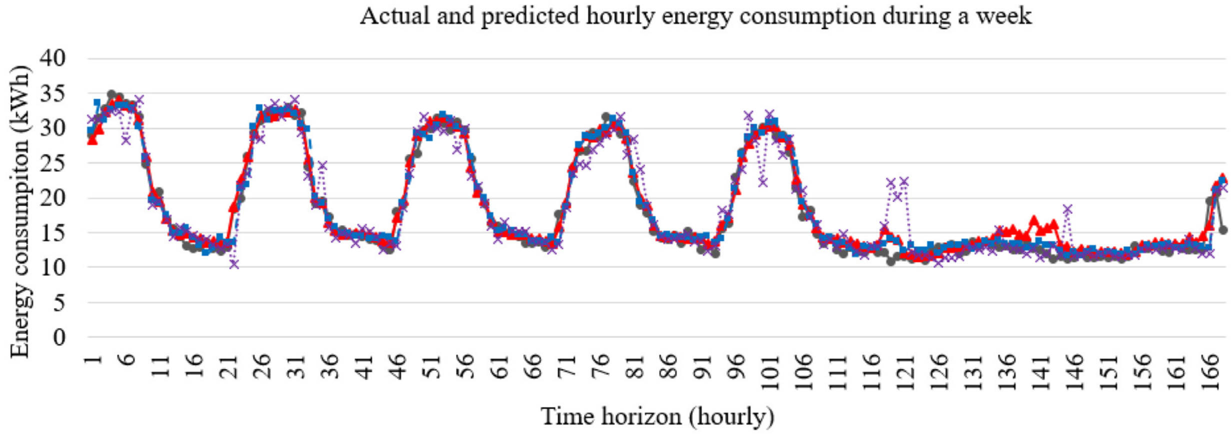


b. Dataset 2

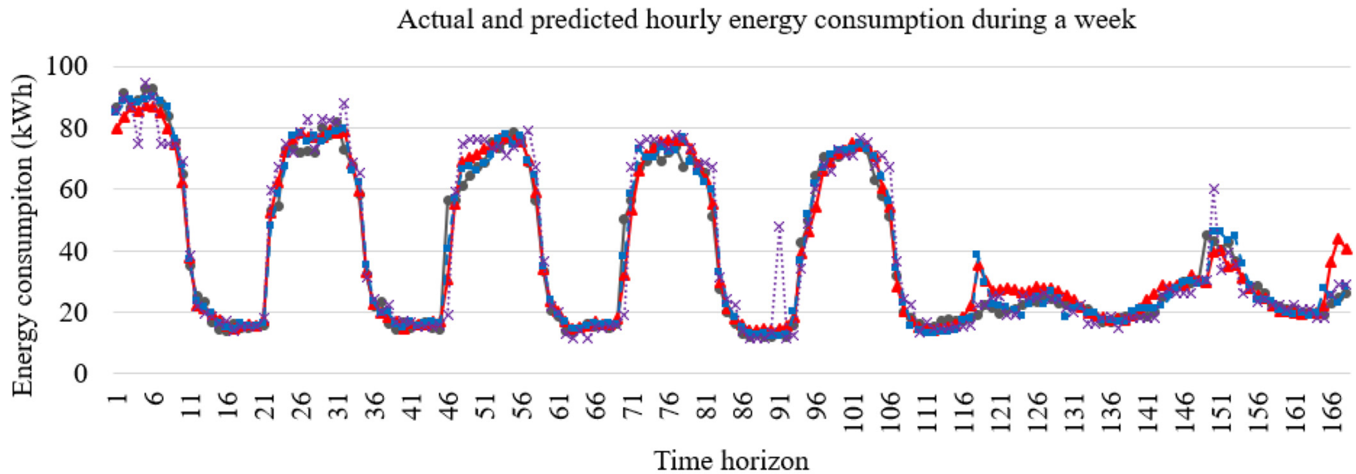


c. Dataset 3

Fig. 7. Actual and predicted energy consumption in building during a week in the test phase.



d. Dataset 4



e. Dataset 5.

Actual RF M5P RT

Fig. 7. (*continued*).

the 1-step-ahead prediction, from 0.612 to 0.940 kWh for the 12-steps-ahead prediction, and from 0.626 to 0.868 kWh for the 24-steps-ahead prediction. The prediction results confirmed that there was a good agreement between predicted and actual data. Prediction results of short-term energy consumption in buildings can support users to act further actions to reduce energy use and improve energy efficiency.

The comparison results presented that the RF model outperformed the M5P and RT models in terms of all performance measures (*i.e.*, MAE, MAPE, RMSE, SI) in predicting the hourly energy use in buildings. For example, for dataset 2, the RF model obtained the smallest MAPE values of 6.023% in the 1-step-ahead prediction of building energy use, followed by the M5P model with the MAPE of 6.331%. Performance of the RF model in forecasting 1-step-ahead building energy consumption was better about 49.21%, 46.93%, and 69.15% in MAE, MAPE, and RMSE than that of the RT model. To sum up, the proposed RF model was an effective prediction model among the investigated ML models. The RF model was stable and robust in predictions that can improve its reliability for building owners in planning their energy-saving strategy.

As the contributions, this study contributes to (i) the state of the knowledge by examining the generalization and effectiveness of ML models in predicting building energy consumption patterns over the operational stage; and (ii) the state of practice by proposing an effective tool to help the building owners and facility managers in understanding building energy performance for enhancing the energy efficiency in buildings. A limitation of this study, parameters of the RF models were set as defaults. The RF's parameters could be fine-tuned by using optimization algorithms to improve predictive accuracy.

In this study, future energy consumption data were predicted based on only this historical energy data. Future work may consider metadata in the prediction. Particularly, the weather condition such as outdoor temperature, building operational schedule, time index (day of the week, an hour of the day) may influence potentially the performance of the prediction models. The seasonality of data should be considered in predictions in future works.

Declaration of competing interest

None.

CRediT authorship contribution statement

Anh-Duc Pham: Conceptualization, Formal analysis, Resources, Funding acquisition. **Ngoc-Tri Ngo:** Writing - original draft, Conceptualization, Methodology, Validation, Project administration. **Thi Thu Ha Truong:** Writing - review & editing, Conceptualization. **Nhat-To Huynh:** Investigation, Writing - review & editing. **Ngoc-Son Truong:** Visualization.

Acknowledgements

Research is supported by Vingroup Innovation Foundation (VINIF) in project code VINIF.2019.DA05.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2020.121082.

References

Abdelkader, S.S., Grolinger, K., Capretz, M.A.M., 2015. Predicting energy demand peak using M5 model trees. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 509–514.
Allouhi, A., El Fouih, Y., Kousksou, T., Jamil, A., Zeraouli, Y., Mourad, Y., 2015. Energy consumption and efficiency in buildings: current status and future trends. J. Clean. Prod. 109, 118–130.
Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renew. Sustain. Energy Rev. 81, 1192–1205.
Breiman, L.J.M.L., 2001. Random Forests, 45 (1), 5–32.
Chan, J.C.-W., Paelinckx, D., 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Rem. Sens. Environ. 112 (6), 2999–3011.
Chou, J.-S., Bui, D.-K., 2014. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy Build. 82, 437–446.
Fan, C., Xiao, F., Zhao, Y., 2017. A short-term building cooling load prediction method using deep learning algorithms. Appl. Energy 195, 222–233.
Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A., 2013. State of the art in building modelling and energy performances prediction: a review. Renew. Sustain. Energy Rev. 23, 272–288.
Goudarzi, S., Anisi, M.H., Kama, N., Doctor, F., Soleymani, S.A., Sangaiah, A.K., 2019. Predictive modelling of building energy consumption based on a hybrid nature-inspired optimization algorithm. Energy Build. 196, 83–93.
Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., Huang, R., Huang, Y., 2018. Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Appl. Energy 221, 16–27.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11 (1), 10–18.
Jiang, R., Tang, W., Wu, X., Fu, W., 2009. A random forest approach to the detection of epistatic interactions in case-control studies. BMC Bioinf. 10 (Suppl. 1). S65-S65.
Killian, M., Kozek, M., 2016. Ten questions concerning model predictive control for energy efficient buildings. Build. Environ. 105, 403–412.
Kneifel, J., Webb, D., 2016. Predicting energy performance of a net-zero energy building: a statistical approach. Appl. Energy 178, 468–483.
Kontschieder, P., Bulò, S.R., Bischof, H., Pelillo, M., 2011. Structured class-labels in random forests for semantic image labelling. In: 2011 International Conference on Computer Vision, pp. 2190–2197.
McNeil, M.A., Karali, N., Letschert, V., 2019. Forecasting Indonesia's electricity load through 2030 and peak demand reductions from appliance and lighting efficiency. Energy Sustain. Dev. 49, 65–77.
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.
Ocampo Batlle, E.A., Escobar Palacio, J.C., Silva Lora, E.E., Martínez Reyes, A.M., Melian Moreno, M., Morejón, M.B., 2020. A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: case study, Federal University of Itajubá (UNIFEI). J. Clean. Prod. 244, 118551.
Qiao, R., Liu, T., 2020. Impact of building greening on building energy consumption: a quantitative computational approach. J. Clean. Prod. 246, 119020.
Qiu, X., Zhang, L., Nagaratnam Suganthan, P., Amaratunga, G.A.J., 2017. Oblique random forest ensemble via Least Square Estimation for time series forecasting. Inf. Sci. 420, 249–262.
Quinlan, J.R., 1992. Learning with continuous classes. In: 5th Australian Joint Conference on Artificial Intelligence, pp. 343–348.
Rafsanjani, H.N., Ahn, C.R., Chen, J., 2018. Linking building energy consumption with occupants' energy-consuming behaviors in commercial buildings: non-intrusive occupant load monitoring (NIOLM). Energy Build. 172, 317–327.
Singh, P., Dwivedi, P., 2018. Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem. Appl. Energy 217, 537–549.
Tian, C., Li, C., Zhang, G., Lv, Y., 2019. Data driven parallel prediction of building energy consumption using generative adversarial nets. Energy Build. 186, 230–243.
Wang, Z., Wang, Y., Zeng, R., Srinivasan, R.S., Ahrentzen, S., 2018. Random Forest based hourly building energy prediction. Energy Build. 171, 11–25.
Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2007. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14 (1), 1–37.
Xie, Q., Ni, J.-Q., Bao, J., Su, Z., 2019. A thermal environmental model for indoor air temperature prediction and energy consumption in pig building. Build. Environ. 161, 106238.
Zhu, Y., Xu, W., Luo, G., Wang, H., Yang, J., Lu, W., 2020. Random Forest enhancement using improved Artificial Fish Swarm for the medial knee contact force prediction. Artif. Intell. Med. 103, 101811.