

Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings



Shalika Walker^{a,*}, Waqas Khan^a, Katarina Katic^a, Wim Maassen^{a,b}, Wim Zeiler^a

^a Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, the Netherlands

^b Royal HaskoningDHV, George Hintzenweg 85, 3009 AM Rotterdam, the Netherlands

ARTICLE INFO

Article history:

Received 26 June 2019

Revised 3 December 2019

Accepted 14 December 2019

Available online 15 December 2019

Keywords:

Hourly demand prediction

Regression-trees

ANN

Groups of buildings

ABSTRACT

As with many other sectors, to improve the energy performance and energy neutrality requirements of individual buildings and groups of buildings, built environment is also making use of machine learning for improved energy demand predictions. The goal of achieving energy neutrality through maximized use of on-site produced renewable energy and attaining optimal level of energy performance at building-cluster level requires reliable short term (resolution shorter than one day) energy demand predictions. However, the prediction and analysis of the energy performance of buildings is still focused on the individual building level and not on small neighborhood scale or building clusters.

In a smart grid context, to better understand electricity consumption at different spatial levels, prediction should be at both individual as well as at building-cluster levels, especially for neighborhoods with definite boundaries (such as universities, hospitals).

Therefore, in this paper, using data from 47 commercial buildings, a number of machine learning algorithms were evaluated to predict the electricity demand at individual building level and aggregated level in hourly intervals. Predicting at hourly granularity is important to understand short-term dynamics, yet most of the neighborhood scale studies are limited to yearly, monthly, weekly, or daily data resolutions. Two years of data were used in training the model and the prediction was performed using another year of untrained data. Learning algorithms such as; boosted-tree, random forest, SVM-linear, quadratic, cubic, fine-Gaussian as well as ANN were all analysed and tested for predicting the electricity demand of individual and groups of buildings. The results showed that boosted-tree, random forest, and ANN provided the best outcomes for prediction at hourly granularity when metrics such as computational time and error accuracy are compared.

© 2019 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license.
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Growing concern about energy and environmental problems have driven interest in the buildings because buildings account for around 40% of the final energy consumption [1,2]. Therefore, enhancement of energy efficiency specifically in commercial buildings has become a non-trivial requirement [3]. In addition, growing penetration of building integrated sustainable energy sources and energy conversion technologies require flexibility of energy systems [4]. However, it is yet to be recognized whether the current

sustainable developments actually lead to real savings of final energy consumption and enhancement of energy flexibility. Therefore, associated regulations have presented targets for buildings and has led to the promotion of the concept of nearly/net-zero energy buildings (nZEB) [5,6].

A nZEB typically contains cohesive renewable energy sources. Usually by adding intelligent control strategies the energy flows of nZEBs are optimally governed to achieve the required user comfort [7,8], and load matching. Load matching is performed typically between on-site renewable generation, grid energy consumption, and energy demand [5], with the aim of consuming the generated on-site renewable energy maximally. Nevertheless, a considerable amount of the generated renewable energy is sent back to the grid without using locally during the spring/summer seasons.

* Corresponding author.

E-mail address: S.W.Walker@tue.nl (S. Walker).

To minimize the associated grid stress with high penetration of renewable energy sources, it is essential to utilize the on-site produced energy locally. This cannot be optimally achieved by existing control strategies, which merely focus on single-building level performance [5]. Therefore, moving the boundaries from single building level to building cluster level is considered important to maximize the local utilization of on-site produced energy and to increase the suppleness of non-dispatchable energy sources using energy storage systems. Effective collaborative controls between neighboring buildings can pave the way towards zero energy neighborhoods with potential energy performance improvement of all associated buildings whether nZEB or non-ZEBs [5]. Thus, smarter use of technologies and available renewable resources is marked essential in achieving maximum techno-economic and environmental benefits [9].

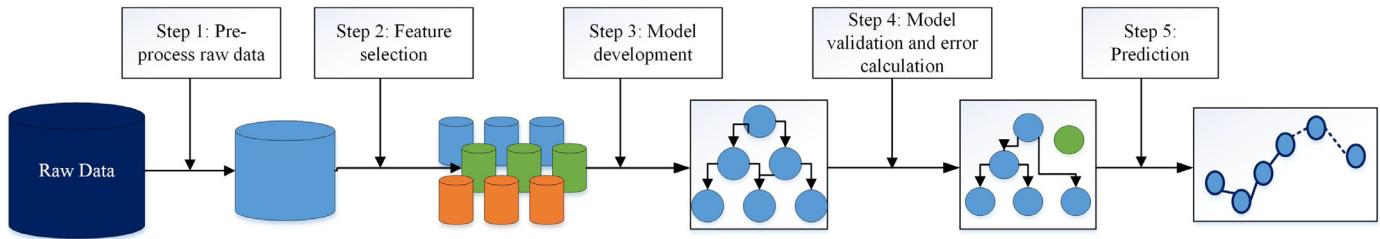
To dispatch the available resources effectively and to identify the overall energy saving potential, short-term and long-term estimation of building energy demand is considered essential. Continuous monitoring, management, and demand prediction also serve the purpose of providing performance targets to building owners and facility managers. Prediction of energy demand can either be classical, using building simulation tools, or data-driven using measurement data [2,4]. Engineering calculation and simulation model-based benchmarking categorize under classical modeling and develops utilizing the physics of buildings. Classical modeling requires a certain amount of building parameters to acquire a reliable demand prediction [6]. On the other hand, given sufficient quantity of historical data [10], statistical modeling and machine learning (ML) algorithms which can be categorized as data-driven modeling [6] could predict future demand reliably without an enormous amount of simulations [6,11]. Therefore, recently, a trend has been developed in research towards estimating building energy consumption using data-driven and time-series analysis techniques [12]. In data-driven analysis, energy demand prediction is performed by employing historical energy consumption data of buildings and other associated parameters such as climate data through a learning process [11]. Regression methods, artificial neural network models (ANNs), etc. [11] can be categorized under data-driven modeling algorithms for such learning processes.

In this study, for electricity demand prediction, ML techniques are used with time-series data. A time-series is a sequence of observations recorded over equal time intervals [13]. For building energy demand forecast, the time-series data can be either yearly, monthly, weekly, daily, hourly, or even smaller data resolutions such as 15 min. Regression analysis for demand prediction can be named as a common denominator for a reasonable number of research papers [1,3,6,11,14,15]. Deb et al. [11] presented nine widely used time-series prediction techniques for building energy consumption including the most extensively used ML algorithms namely ANN and support vector machine (SVM). Ahmad et al. [15] compared ANN and SVM in their study and concluded it is hard to decide which algorithm forecasts the best. Moving towards more advanced algorithms, Mocanu et al. [16] used deep learning auto-regressive estimation methods for the prediction of building electricity consumption. The authors concluded that the prediction could be improved by using extra model inputs, such as outdoor temperature and time information (Month, Day). Several other studies [16–21] also discussed the accuracy of different ML algorithms in demand prediction at the building level for a certain prediction horizon. Out of them, some studies [16–20] emphasized that ML algorithms in their basic form are difficult to use for demand prediction. Therefore, modifying the ML algorithms from their basic structure [16–19] has also been tried in literature to improve the energy demand prediction and control [21] capability of the buildings. However, most of these existing literature has focused only on single building level.

In the future, when smart grids would be in operation, it is unlikely that the prediction and management of energy demand at single building level will be a true objective of the network operators. Instead, the aggregated demand of building clusters will be vital in order to improve the energy performance and to let all buildings collaborate with each other, while cleverly exploiting each other's peaks and valleys of the demand profiles. Thereby, achieving nearly constant demand behaviours and ultimately energy neutrality at neighborhood level.

Within the context of predicting multi-building energy use patterns, Xu et al. [20] used ANN to predict energy demand by using prototype building models. Because of the difficulty of collecting historical data of all the buildings, the authors have predicted monthly electricity consumption of reference buildings first and tried to estimate the correlation of the reference and non-reference buildings separately. Using the data collected by the US Department of Energy's building performance database (BDP), Robinson et al. [3] attempted to predict the city scale annual energy consumption of commercial buildings using ML algorithms and a limited set of building-level features. Their paper emphasized the importance of predicting in aggregated building scale to avoid the prediction errors of individual buildings and to obtain useful insights into cities/neighborhoods. Using a similar database, but for both commercial and residential buildings of New York City, Kontokosta et al. [22] predicted the annual energy consumption using three ML algorithms. The authors emphasized citywide energy prediction is crucial in understanding carbon reduction measures. However, these research-work which attempted to estimate energy consumption at the neighborhood/city scale were mostly focused on long-term energy prediction (monthly, yearly). For a future smart grid operation, to optimize load scheduling [23] and to achieve the maximum benefits and collaboration between buildings, long-term demand prediction would not be advantageous as of short-term demand prediction [24]. Using the New York Independent System Operator's (NYISO) electrical load data set, Ahmed et al. [24] conducted a study that used regression trees for demand prediction at city level. In their study, the authors highlighted the importance of short-term day-ahead load prediction for all days of the calendar year. Another study [25] tried to compare deep learning and machine learning algorithms at an aggregated level for the Dutch electricity market using 15-minute demand resolutions. This study argues that including influencing factors such as pricing does not necessarily improve the accuracy of prediction at aggregated level, because of the growing uncertainties in the electricity-grid demand portfolio due to predominant role of solar energy.

The aforementioned factors emphasized the necessity of reliable short term energy demand prediction of the buildings. Reliable prediction is needed for achieving energy neutrality through maximized use of on-site produced renewable energy and attaining an optimal level of energy performance. Prediction in a localized context is essential to understand the energy consumption and optimization of parameters to improve energy efficiency. On the other hand, aggregated demand prediction is crucial for optimization of the market-flows and economic feasibility (Ex: computational cost). Therefore, demand prediction assessment of buildings should be at different spacial levels, which are otherwise missing in the existing literature. This type of analysis could certainly benefit neighborhoods with an already specified boundary such as hospitals or campuses. Therefore, to fill the void, in this paper, prediction of the electrical energy consumption of commercial buildings using different ML algorithms is performed locally and at building-cluster level at hourly granularity. In addition, the performance accuracy of these algorithms was compared using individual scale predictions and aggregated scale predictions. It was determined in different spatial levels the type of information needed to get an improved forecast.

**Fig. 1.** Demand prediction procedure.**Fig. 2.** Case study campus.

2. Method

In this study, the energy demand prediction procedure using several prediction methods was established following the five steps shown in Fig. 1;

- Step 1: Pre-processing of raw data
- Step 2: Feature selection
- Step 3: Model development
- Step 4: Model validation and error calculation
- Step 5: Prediction

2.1. Description of the raw dataset

This analysis is performed using actual hourly electricity consumption data of a campus together with the attributes on building-level. The case study campus is located in the Netherlands and mainly addresses technology companies. The campus consists of 47 buildings with multiple buildings originating since the foundation in 1968. The campus as a whole (not building by building) is planning to become energy neutral before the year 2030. With this vision of becoming an energy neutral neighbourhood on its own, energy demand prediction is a crucial matter to this campus to identify the possible renewable energy initiatives. The geographical configuration of the campus buildings is presented in Fig. 2, and the overview of the existing buildings is shown in Table 1.

In Table 1, the total number of 47 buildings are categorized according to their functionality; namely, office buildings, industrial buildings, and other functions. Under each of these categories, the buildings are further divided according to their year of construction. The representation is given using the total floor area of the buildings. During the raw data collected period from 2016 to 2018,

Table 1
Overview of the characteristics of the buildings.

Building Type	Year built	Floor area (m ²)	Total (m ²)
Office buildings	1968	126,454	249,800
	1991	18,612	
	1998	1361	
	2005	39,202	
	2006	19,015	
	2007	18,665	
	2009	20,144	
	2013	1309	
	2014	110	
	2015	1928	
Industry buildings (ICT/Process/Storage/ Cafeteria)	1968	35,674	66,186
	1998	11,272	
	2005	11,188	
	2014	4894	
	2015	3158	
Other buildings	1968	190	1570
	2008	273	
	2009	1107	

no new buildings have been added to the campus. It was observed that about five buildings were completely renovated or had a significant change in function. However, exact information about the renovation or changes in the functionality of the buildings couldn't be collected from the building management.

A data acquisition and control system monitors the operation of the case study campus. Hourly operational data for electricity was collected from this system from January 2016 to December 2018. Weather data corresponding to this period were acquired from the KNMI (Dutch Royal Meteorological Institute) weather station 328, which is the nearest to the campus. Having full data for more

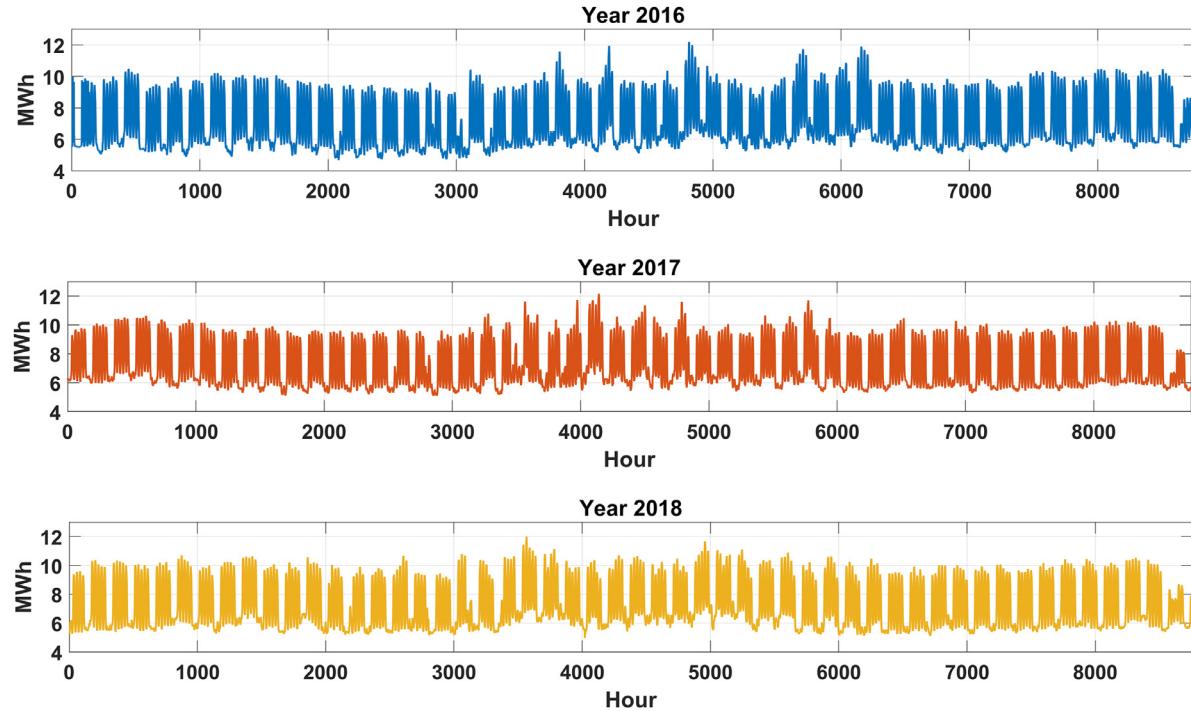


Fig. 3. Hourly aggregated level electricity consumption for 2016–2018.

years provide a better prediction model, which is described under Section 2.4. Fig. 3 illustrates the aggregated electricity consumption of the entire neighbourhood in hourly intervals for the years 2016 to 2018.

2.2. Step 1: pre-processing of raw data

Prior to the development of the model, the collected data were preprocessed to identify whether it is needed to be cleaned. The cleaning of data was based on two conditions, which are namely erroneous and missing data points. However, no missing data points were identified in the data set. For this study, Lagrange polynomial interpolation is used to eliminate the erroneous data points [20]. Errors can be occurred due to instrument malfunction and equipment faults [1]. If the time series can be presented as $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, the Lagrange interpolation is formulated as presented in Eq. (1).

$$L(x) = \sum_{i=0}^n y_i \prod_{j=0, j \neq i}^n \frac{(x - x_j)}{(x_i - x_j)} \quad (1)$$

In the equation, $L(x)$ is the Lagrange polynomial, y represent the interpolation value, and n is the size of the data used for interpolation. By visual observation, only a few number of erroneous outliers were identified. Then, by using the Lagrange interpolation, the erroneous values are censored using 150 data points before and after the erroneous outliers. In the outcome after the interpolation, only the identified few erroneous data points were altered.

2.3. Step 2: feature selection

After the data set is pre-processed, the next step is to identify the most important inputs or features that will affect the forecast. A feature is a variable that contains relevant information in predicting the output [1,20].

Based on common knowledge and understanding of building operation, weather-related variables are considered significant as

features. As weather features, outdoor temperature, dry bulb temperature, and relative humidity (RH) were chosen. Prediction of energy use of commercial buildings is difficult and complicated, therefore, other than the weather parameters, day of the week, hour of the day, month of the year, and seasons have been used as categorical features. One intervention event which is “working day”, taking only the values 1 and 0 is also considered as a feature [13]. Other than that, in order to improve the prediction accuracy, energy consumption related autoregressive parameters have been used; namely the energy consumption of the previous day and energy consumption of the previous week. The aim is to obtain a day-ahead prediction; therefore, energy consumption of previous hours is not of interest. Feature selection has a non-trivial effect on the accuracy of the model [22]. Consequently, it is essential to choose adequate yet, less number of features. By choosing a less number of features, the model becomes more general and straightforward [3], which helps in applying it to similar case studies.

2.4. Step 3: model development

The model development aims to estimate the energy consumption of the case-study cluster of buildings in hourly intervals. This objective is expressed in a machine learning algorithm as follows. Given the \mathbf{X} features and past energy consumption data \mathbf{Y} Eqs. (2) and (3), the target energy consumption is predicted through a trained model $f(\mathbf{X})$, as illustrated in Fig. 4.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ \dots & & & \\ x_{1t} & x_{2t} & \dots & x_{nt} \end{bmatrix} \quad (2)$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \dots \\ y_t \end{bmatrix} \quad (3)$$

In the matrix \mathbf{X} , ‘1 to n’ columns represent the predictors and ‘1 to t’ rows represent the time-steps in the training data set. The vector of responses \mathbf{Y} presents the same number of observations as

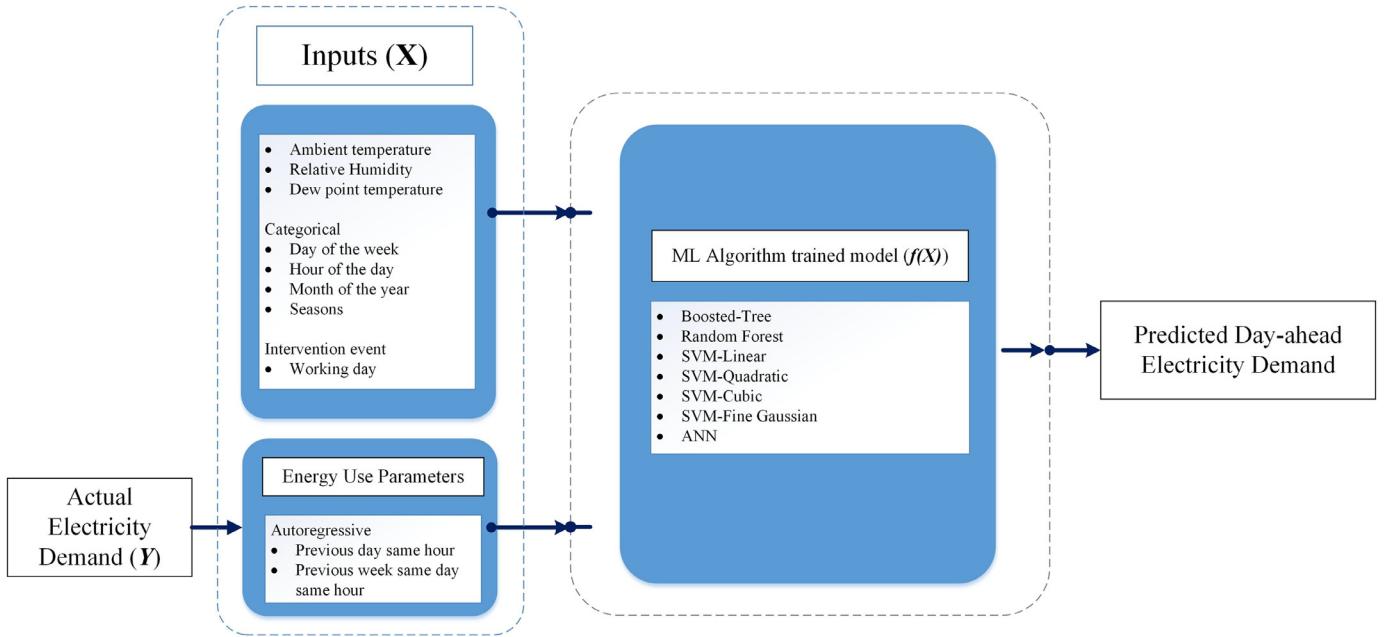


Fig. 4. Model development.

Table 2
ML algorithms used in literature.

Algorithm	Reference
Boosted-tree/Random Forest	[27,28,14,3,22,29]
SVM	[30,28,15,14,3,22,12]
ANN	[27,30,31,32,33,15,12]

the rows in \mathbf{X} . Models were developed using the machine learning toolbox of MATLAB 2017b [26].

Since the selected case study buildings are commercial buildings, it is a possibility that the functionality of the buildings differs throughout the years. Thereby, the scales of the dataset can vary. In that case, it is important to pay attention and standardize the collected past energy consumption data. Standardization is the process of rescaling the data set when there are differences in scales of magnitude such that the statistical properties remain approximately constant over time. From observation, it was distinguished, data collected from 12 buildings needed to be standardized. The standardization process rescales the distribution such that the mean of the observed values is set to 0 and the standard deviation is set to 1. In MATLAB black-box modeling the standardization process can be performed.

To develop the ML models $f(\mathbf{X})$ and to compare the performance, following algorithms have been used. In general, these models attempt to adjust the internal parameters according to the given features and past energy consumption data, and minimize a loss function between the target values and values predicted [3]. The following ML algorithms have been chosen mainly because of the extensive usage and performance superiority according to the literature; see Table 2.

2.4.1. Regression decision trees

A decision tree is a non-parametric approach that identifies different ways of splitting a data set based on conditions until the information gain is zero. Construction of a tree usually inherits a top-down approach where a variable is chosen at each step, which 'best' splits the set of data [34]. The 'best' split is distinctive for the algorithm used. In order to predict responses, decisions are fol-

lowed from the root node to the leaf nodes [34]. Generally, there are two steps to building a regression tree.

- 1 The set of possible values of the predictors, also known as predictor space (\mathbf{X}) is divided into J distinct non-overlapping regions R_1, R_2, \dots, R_J . The regions are constructed in a way that it minimizes the residual sum of squares given by Eq. (4):

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \bar{y}_{R_j})^2 \quad (4)$$

The mean response of the training observations within the j^{th} region is given by \bar{y}_{R_j} .

- 2 Then, for every new observation that falls in region R_j , the prediction gives the mean of the response values for the training observations in R_j .

The conventional regression tree algorithms suffer from overfitting and large variance [34]. Therefore, some techniques combine more than one decision tree in order to avoid these major drawbacks of conventional decision trees. These algorithms are called **Ensemble methods**, where a group of weak learners is combined to form a strong ensemble which then produces better predictive performance [34,35]. In this study, two ensemble regression trees namely, the boosted tree and random forest are used for the demand prediction. Random forest is an improvement of the original bagged tree with a small tweak.

Generally, bag constructs deep grown trees while boost algorithms create shallow trees. Thereby, bagged trees lead to relatively slow predictions than boost trees. In bagging, the training data set is bootstrapped by taking repeated samples. Thus, B number of different bootstrapped samples are generated. The prediction is performed as discussed in conventional decision trees for all these samples, and finally, they are averaged to obtain a single prediction as given in Eq. (5). $\hat{f}_i(x)$ represents the prediction of the i^{th} bootstrap sample. The effectiveness of the prediction is said to be higher in bagged trees [35].

$$f(x) \leftarrow \frac{1}{B} \sum_{i=1}^B \hat{f}_i(x) \quad (5)$$

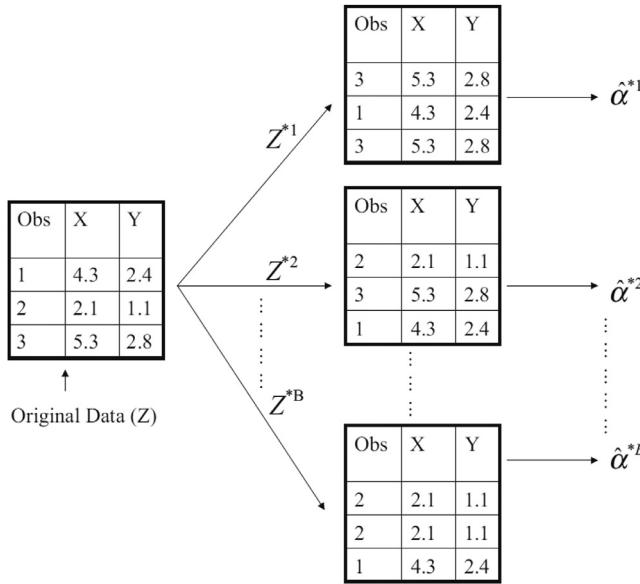


Fig. 5. Illustration of bootstrap as per reference [34].

Graphical illustration of bootstrap on a small sample containing $n = 3$ observations is presented in Fig. 5. In the figure, α represent the estimate of each bootstrap data set.

In random forest, when the training samples are bootstrapped and a number of decision trees are made, a random sample of m predictors are used out of the full set of X predictors, and each split in a tree is allowed to use only a subset of the m predictors. This process makes the prediction more reliable. Unlike bagging or random forest, which involves creating multiple copies of the original training data set and fitting a separate decision tree to each copy, boosting works with sequentially grown trees [34]. Here, each tree is grown with the information obtained from previously grown trees. Thereby, each tree is fit on a modified version of the original data set, as shown by Eq. (6). Parameter λ is a small positive number that controls the rate at which the boosting learns and B represents the number of trees [34].

$$f(x) \leftarrow \sum_{b=1}^B \lambda \hat{f}_b(x) \quad (6)$$

2.4.2. Support vector machine (SVM)

SVM is a robust learning algorithm for solving non-linear problems and can be used for both regression and classification [8]. SVM is used to find an optimal hyperplane that separates the classes with a maximum margin [34]. If the SVM is exercised to predict a time series or real numbers, it is called support vector regression (SVR). SVR uses the same principles as SVM. For every input parameter vector (\mathbf{X}) and its corresponding output vector (\mathbf{Y}), SVR relates the inputs and outputs using Eq. (7) [6].

$$\mathbf{Y} = \mathbf{W} \cdot \varphi(\mathbf{X}) + b \quad (7)$$

\mathbf{W} represents the weight vector, and b represents the bias, which are dependent on the selected kernel function; in this context, linear, quadratic, cubic, and fine-Gaussian. The kernel function quantifies the similarity of two observations [34]. Even though SVM has its advantages such as the ability to train with a fewer number of samples and contains a fewer number of hyperparameters than Artificial Neural Network (ANN) [6], the computation time acts as a significant drawback of this algorithm. In this study, SVM has been used along with regression trees and ANN, and the performances of prediction are compared.

2.4.3. Artificial neural network (ANN)

ANN is a widely used model for the prediction of non-linear, complex problems [12] and is one of the main techniques used for Deep learning. ANN models are used for deriving meaningful information from imprecise and complicated data [11]. An ANN is considered an expert when it is trained in a specific category of data patterns. When using ANN, the neural network is trained on a dataset that consists of mapped pairs of inputs and outputs. During training, each neuron is assigned a numeric weight (W). Together with an activation function, these weights define the output of each neuron. ANN starts learning by adjusting its weights iteratively, so that it can map inputs across outputs and learn from the dataset. Once the network is trained, and the weights are specified for the connection between the neurons, it is used for validation on a new dataset. Every input layer neuron is connected to neurons in a hidden layer, which is connected to the output layer neurons. A complete connection is formed from the input layer to the output layer, as shown in Fig. 6. Each weight is just a factor that changes throughout the whole process until the loss function is minimized [36].

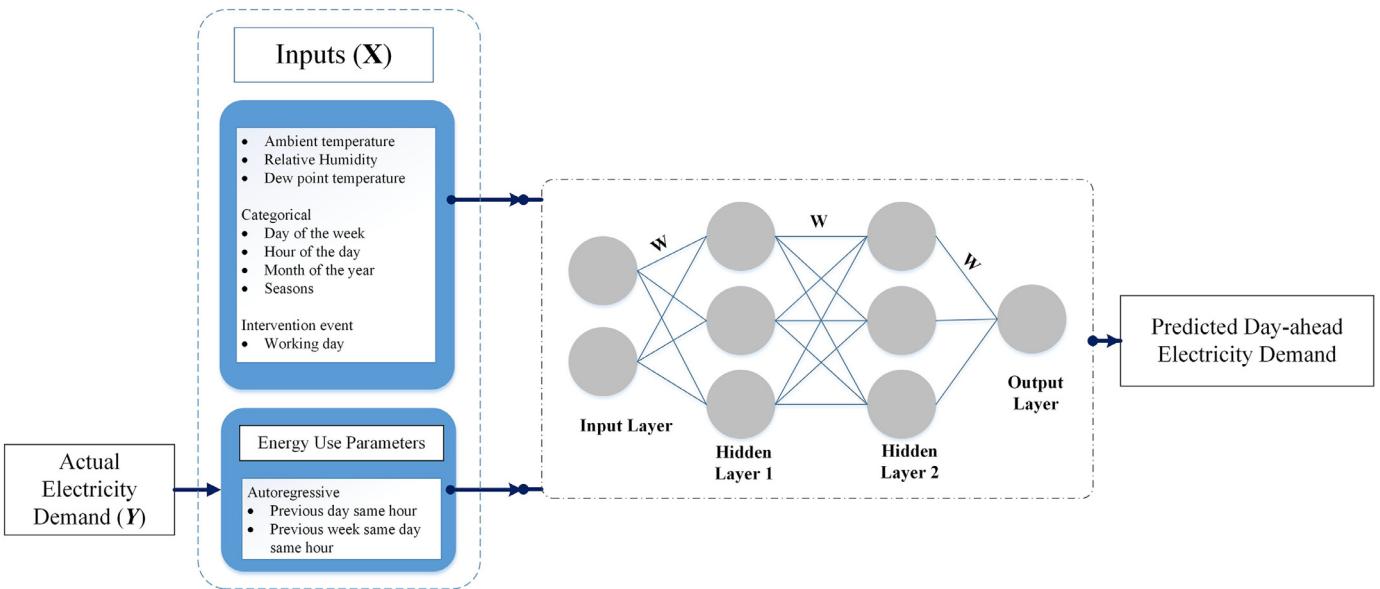


Fig. 6. Simple Artificial Neural Network.

In this paper, a feed-forward neural network was used with a sigmoid activation function [37] for the hidden layers and a linear function at the output layer because it is a regression problem. Sigmoid function was used at the hidden layers to introduce non-linearity to the network, for learning and allowing the algorithm to generate complex mappings between the features and the outputs, which are necessary for forecasting complex data [38]. Out of the several methods available for training a neural network, such as Error backpropagation, Gauss-Newton, and Levenberg Marquardt backpropagation (LVB), the latter was chosen [39] as the training method of the neural network of this paper. This is because of its faster computation capability [40].

2.5. Step 4: model validation and error calculation

For the validation process, data collected from years 2016 and 2017 (training dataset) were used. As the validation method for regression decision-trees and SVM trained models, k -fold cross-validation was executed [34]. In k -fold validation, the set of training data is divided into k -groups of approximately equal size. In this study, the k value is equal to 10; thereby, 10-fold cross-validation is performed. In each iteration, one group out of the k -groups is treated as a validation set, and the rest of the $k-1$ groups are used to create the model. The held-out validation group is then predicted using the created model. Likewise, the same procedure is repeated k -times, and in each iteration, a different group is picked as a validation set. For ANN, the validation is performed by dividing the training dataset into two parts. From the training data set, randomly chosen 30% of data is held without using to create the model. Then, this set of data is used to validate and test the established model. After model creation and validation based on 2016–2017 data, the year 2018 new dataset was used for demand prediction.

The accuracy of the trained model (with 2016–2017 data) and prediction (test-error) using the unseen year 2018 data were calculated separately. The below-mentioned error performance matrices (2.5.1 – 2.5.4) were used to assess the reliability of all prediction models. In general, for multi-step ahead forecasting as in this study, the relative forecasting errors (MAPE) seen at the aggregated level has been quite low (up to 5%) [41,42] while the forecasting performance at individual level is seen to be much higher (up to 40%) [43,44].

2.5.1. Mean absolute percentage error (MAPE)

This indicator shows (Eq. (8)) the mean percentage error between the predicted (\hat{y}) and the actual (y) energy demand of the buildings on the sample size N . If the MAPE value is less than 10%, it is considered as highly accurate, while 11–20% is regarded as a reasonable forecast [45]. MAPE has been used as a mean of comparison between several algorithms [46] for electricity consumption predictions.

$$\text{MAPE} = 100 * \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{|y_i|} \quad (8)$$

2.5.2. Coefficient of variance of the root mean square error (CV-RMSE)

'ASHRAE Guideline for measurement of energy demand' presents CV-RMSE as a suitable performance calculation matrix for engineering applications. CV-RMSE represents the ratio of Root Mean Square Error (RMSE) and mean of the observations (Eq. (9)). RMSE; Eq. (10) presents the magnitude of the estimation error. However, RMSE is dependent on the sample size (N) and the scale of observations [14]. This means that the RMSE by itself is not informative about the precision of the estimator without information about scale and sample size. Therefore, it is divided by the mean of

the data; Eq. (11) [47]. Then it becomes a scale-independent coefficient. It is specified in ASHRAE guideline, for hourly data resolution predictions, a result with a CV-RMSE value below 0.3 (=30%) [14,48] is sufficiently close to physical reality and adequate [49] for engineering purposes.

$$\text{CV - RMSE} = \frac{\text{RMSE}}{\text{Mean (observations)}} \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (10)$$

$$\text{Mean} = \frac{\sum_{i=1}^N y_i}{N} \quad (11)$$

2.5.3. Coefficient of determination (R^2)

R^2 -squared represents the predictable proportion of the variance in the dependent variable that can be expounded by the independent variable(s), which is presented by Eq. (12). Closer the R^2 -squared value to one, better the model performance. It is taken as a measure of validity for the models because it is a widely used [1,3,48] indicator in statistical analysis. In the equation, \hat{y} -hat and y present the predicted, and the actual energy demands and \bar{y} -bar shows the mean of the real observation set.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

2.5.4. Theil U-Statistics

This indicator is used for forecast accuracy and forecast quality estimation [50]. $i + +x$ in Eq. (13) represents the prediction (\hat{y} -hat) or actual (y) energy demand of time $i + +x$. The indicator produces values greater than zero. If $U < 1$, the forecast is better than the naïve method (Naïve forecasting is using the last period's actual value as the next period's forecast. Since this study is focused on day-ahead prediction, *next period* represents 24 h ahead). Smaller the U -value, better the prediction. If it is $=1$ or >1 , there's no added value of using the proposed forecasting model. In this study, Theil U indicator is used to provide the evidence of predictive capability of the ML algorithms using unseen data.

$$U = \sqrt{\frac{\sum_{i=1}^{N-24} (\hat{y}_{i+24} - y_{i+24})^2}{\sum_{i=1}^{N-24} (\hat{y}_{i+24} - \bar{y})^2}} \quad (13)$$

3. Results and assessment

The results are discussed in two separate sections; individual and collective. In the individual building assessment, each building is analyzed separately, and the performance of the training and prediction models for each building is discussed. In the collective case, all buildings have been used as one group and the performance is discussed accordingly. For the analyzed different ML algorithms, the training and prediction results are presented following the steps introduced in Section 2. During training, for hyper-parameters the *general-values* shown in Table 3 were used. These general-values were chosen by changing the hyper-parameters manually. In general, the best performing combination of hyper-parameters for all the buildings is presented by these values.

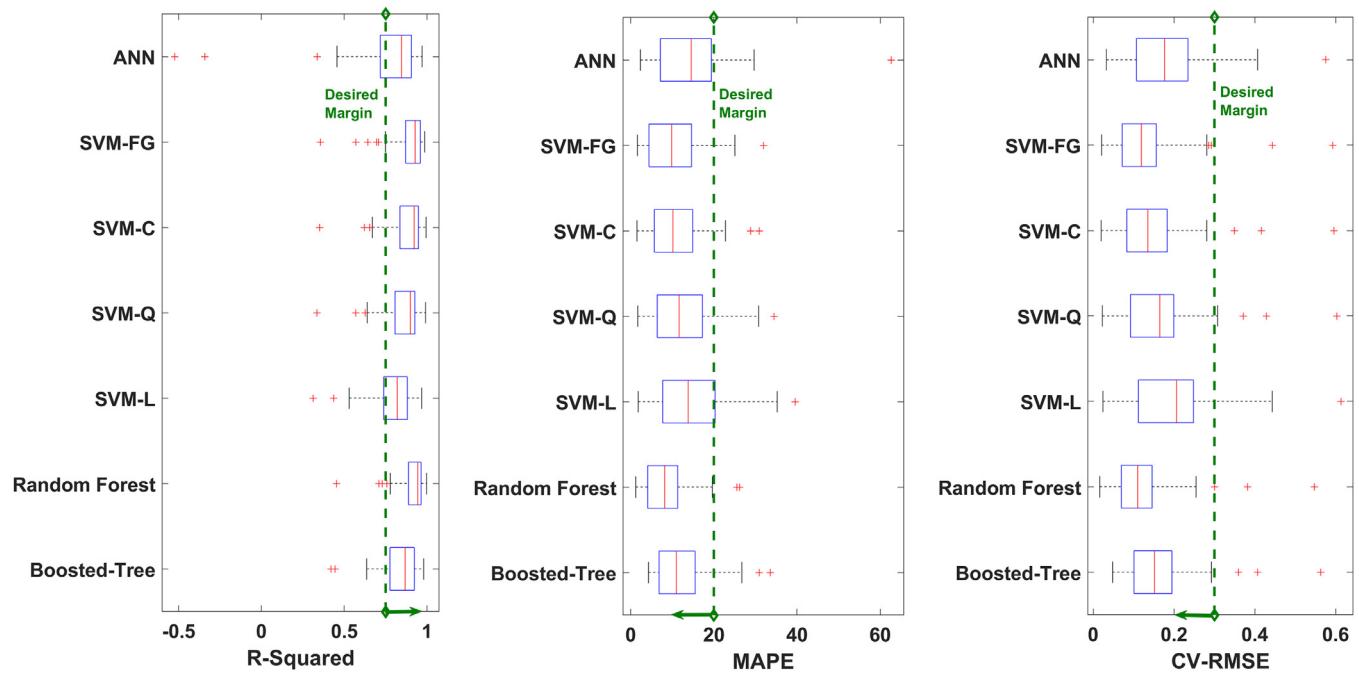


Fig. 7. Deviations of the errors for the 47 buildings.

Table 3

General parameters for all buildings and tuned hyper-parameters for Building-19.

Algorithm	Parameter	General values	Tuned values
Boosted-tree	Minimum leaf size	8	14
	Learning rate	0.1	0.13
	Number of learning cycles	30	133
Random Forest	Minimum leaf size	8	3
	Number of learning cycles	30	500
SVM-Linear (SVM-L)	Box constraint	26.68	917.19
	Epsilon	2.67	0.33
	Kernel scale	36	887.99
SVM-Quadratic (SVM-Q)	Box constraint	26.68	998.93
	Epsilon	2.67	0.23
	Kernel scale	36	20.96
SVM-Cubic (SVM-C)	Box constraint	26.68	0.31
	Epsilon	2.67	0.16
	Kernel scale	36	13.85
SVM-Fine Gaussian (SVM-FG)	Box constraint	26.68	739.86
	Epsilon	2.67	0.26
	Kernel scale	0.83	10.29
ANN	HiddenLayerSize	10	15

For underperforming buildings, the hyper-parameters were tuned by performing grid searches so that the performance of the individual models are improved. Grid search is used to find the optimal parameter combination of a model that gives the most accurate results [51]. The tuned parameters of Building-19 for each model are also shown in Table 3.

Table 4

Individual assessment summary for trained models.

Error	Margin used	Number of buildings							
		Training models							
		Boosted-tree	Random Forest	SVM-L	SVM-Q	SVM-C	SVM-FG	ANN	
MAPE	Less than 20%	41	44	35	40	41	38	36	
CV-RMSE	Less than 0.3	44	44	40	43	44	45	40	
R ²	Higher than 75%	41	44	34	40	42	41	30	

3.1. Individual building assessment

The results associated with the training and validation of ML models using 2016 and 2017 electricity consumption data of each building are discussed in Section 3.1.1.

3.1.1. Error calculations for the trained models – individual buildings

Fig. 7 shows the deviation of errors for each trained model when the buildings were assessed individually. Table 4 represents, in summary, the number of buildings that are satisfying the appropriate error margins for these trained models. From these error calculations, it can be detected that the boosted-tree and random forest perform best along with SVM-cubic model.

In Fig. 7, it can also be observed, few numbers of buildings (see the red '+' signs) do not perform well according to the chosen error matrices. This indicates that the past data patterns of some buildings do not perform in favor of creating a model for prediction regardless of the standardization process.

A zoomed overview of a poorly performing building (Building-19) is presented in Fig. 8. The horizontal dashed-lines in the figure indicate the desired margins.

Knowing the performance, the next step is to choose suitable trained-models for demand prediction using new data. When selecting best-performing algorithms, computational power and time are also considered important. One can argue, computational time can be improved with better machines and this factor is less important and shouldn't be weighted equally with accuracy. However, better and powerful machines always come with an

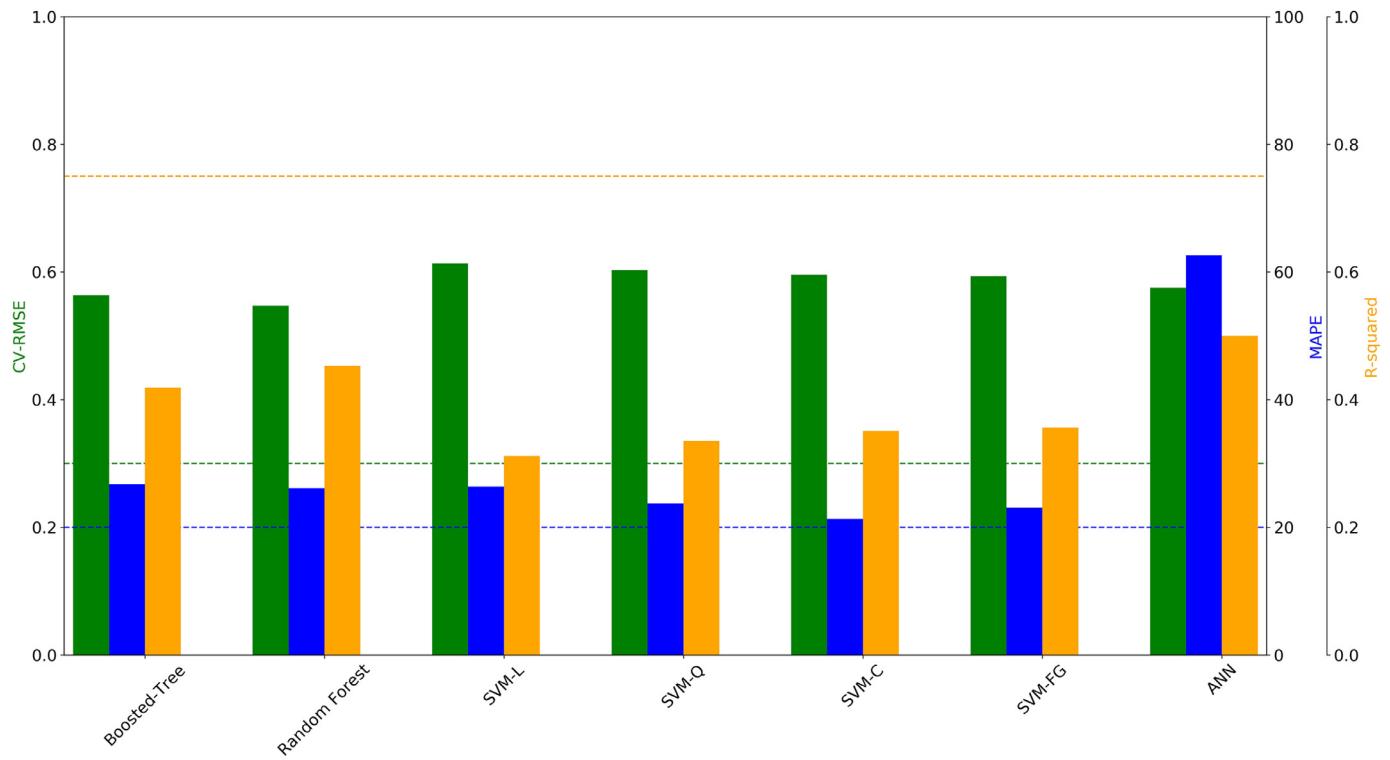


Fig. 8. Error matrices for poorly performing trained models of a building.

additional cost. Selecting ML algorithms giving priority to computational time or accuracy completely depends on the application it will be used. For some applications such as medical requirements, accuracy of the prediction is paramount regardless of the computational time. In this research work, time, and costs are considered important. These factors shape the speed with which the technology will develop within the built environment, and therefore, it helps to define the economic impact. These economic impacts are more than merely the amount of processing power available, but also the details of computational architecture, the actors involved, and the co-evolution of the machine learning field itself [7]. Therefore, computational time needed for training the models has also been calculated.

When comparing the simulation times needed in creating these models, SVM-cubic took more than 17 h to complete the training models of all 47 buildings. In comparison, boosted-tree and random forest completed the task in less than 12 min and ANN in 5 min. According to the performance of the utilized computer, the computational time needed for the creation of training models of all 47 buildings is presented in Table 5.

$$\text{Computational time (CT) total} = \sum_{i=1}^{47} CT_{\text{Building}_i}$$

Considering all these facts, for demand prediction using unseen data, boosted-tree and random forest has been chosen to represent the regression models together with ANN.

3.1.2. Error calculations for the predictions with unseen data – individual buildings

After creating the ML models, the next step is to perform the prediction. For the prediction, new data of the year 2018 is used. Fig. 9, Fig. 10, and Fig. 11 illustrate the MAPE, CV-RMSE, and R^2 value variation of the predictions using the trained models of the 47 buildings. In the three figures, the desired error margin is indicated with a dashed horizontal line. For MAPE, CV-RMSE, and R^2 , the error matrices below 20%, below 0.3 and above 0.75 respectively, are chosen as the desired values as described in Section 2.5.

In the MAPE assessment, with random forest models, 36 buildings stayed under the desired margin, and for boosted-tree and ANN, 35 and 36 buildings fulfilled the margin requirement. Similarly, the CV-RMSE margin was met by 40, 40, and 42 buildings for the boosted-tree, random forest, and ANN, respectively. With R^2 , it can be observed, in all three cases, less than 40 buildings out of the 47 fulfilled the satisfactory margin. In Fig. 11, “undesirable result” indicates a minus value obtained for R^2 .

When comparing the figures, it is possible to see that for some buildings (Ex: Building-30), MAPE and CV-RMSE provide desirable results but, R^2 does not perform well. One possible reason for this could be the non-linear behavior when applying the models to untrained data. R^2 is not considered as a good indicator for such cases [52]. CV-RMSE and MAPE provide better indication of the model's usefulness in such occasions. More insights on the models' overall precision can be obtained with the Theil-U indicator. Fig. 12 shows the calculated Theil-U values. From this indicator, it is possible to say that almost all the prediction models perform better than the

Table 5
Computational time needed for each algorithm.

Algorithm	Boosted-Tree	Random Forest	SVM-L	SVM-Q	SVM-C	SVM-FG	ANN
CT (minutes)	5	11	107	258	1020	173	5

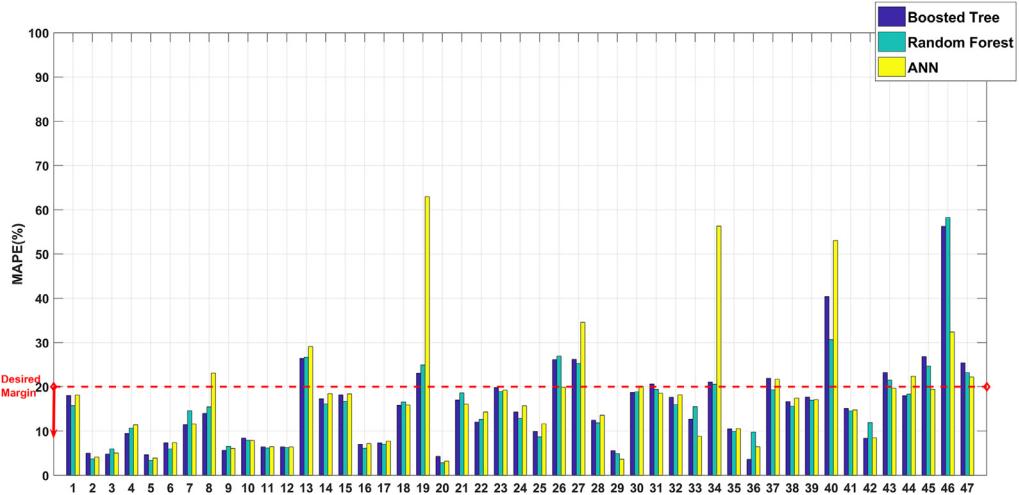


Fig. 9. MAPE calculation for the prediction data set.

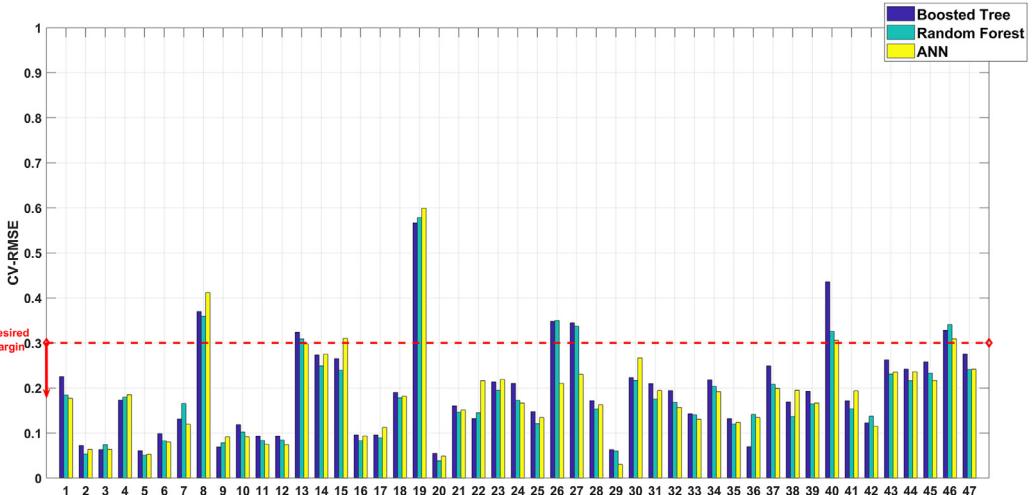
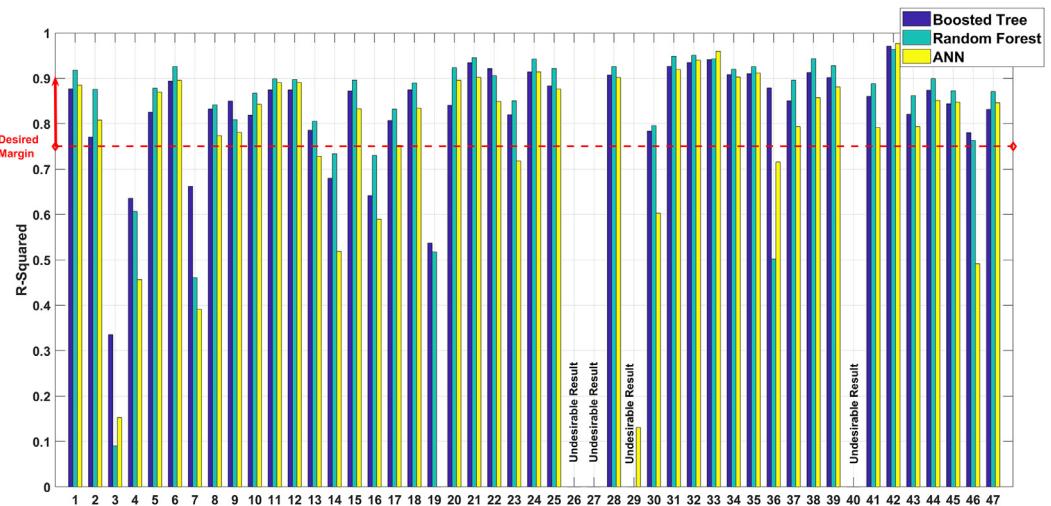


Fig. 10. CV-RMSE estimate for the prediction data set.

Fig. 11. R^2 calculation for the prediction data set.

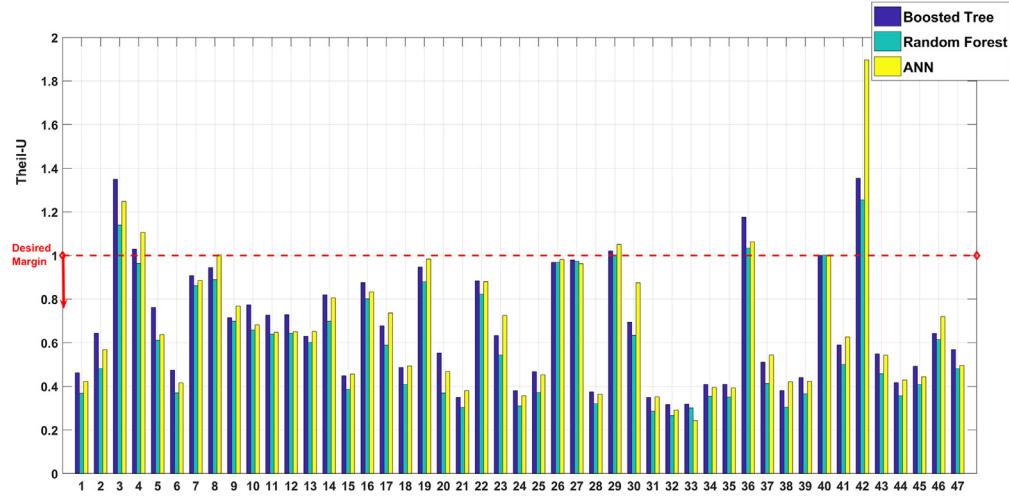


Fig. 12. Theil-U calculation for the prediction data set.

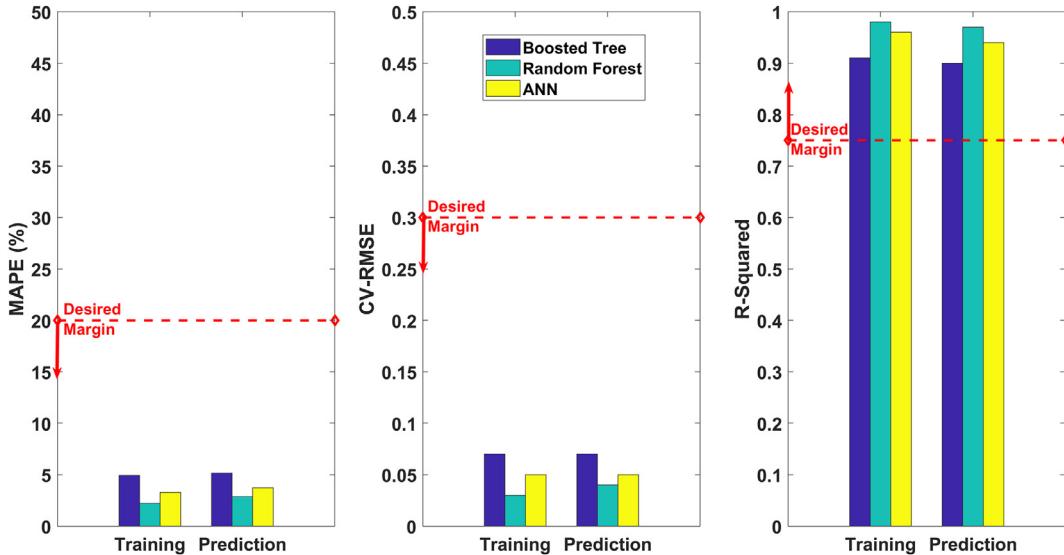


Fig. 13. Error calculation for aggregated-level trained models (2016–2017 data) and prediction (unseen 2018 data).

naïve forecast. For Building-3 and 42, even though the prediction models perform acceptably with MAPE, CV-RMSE, and R^2 calculations, Theil-U indicator conveys naïve estimate is better than the created ML models. On the contrary, regardless of the poor accuracy calculations, the ML model performs better than naïve forecast, for models such as Building-19.

Overall, for the individual building assessment, it is clear from the results for about five buildings the day-ahead electricity consumption cannot be predicted with enough accuracy. One apparent reason for this is the precision of the trained models. Given the fact that the trained models of some buildings did not provide satisfactory outcomes, predicting with new data with enough accuracy was not a possibility. Other than that, given the irregular behavior of electricity demand of a few buildings standardization does not guarantee better predictions. Therefore, it cannot be considered as a smart choice in attempting to make precise predictions for these buildings on an individual scale.

3.2. Accuracy determination of the applied methods on building cluster level

In this section, the trained model accuracy and prediction capability were assessed when these 47 buildings are considered as

a cluster. Even though some buildings presented unpredictable behavior when seeing individually, the effect of these unpredictable patterns can be alleviated to a considerable extent when considering them as groups. Fig. 13 shows the error calculations for the cluster level trained models and predictions. The results show very high accuracy in the training and prediction models. It is also noteworthy to see that the ANN trained model performance is interestingly improved when compared with individual buildings' trained ANN models. The clustered training and prediction significantly reduced the needed time for all models and offered a more reliable load profile for the system. The models were able to train in less than a minute.

The errors of the sum of individual predictions (\hat{Y}) as shown in Eq. (14) also have been calculated, in order to observe the accuracy when compared with the individual predictions (\hat{y}). Fig. 14 below illustrates the resulted error matrices. Even in this case, it is possible to see that the effect of unpredictable data sets has been died out by grouping them as one prediction.

$$\hat{Y} = \sum_{i=1}^{47} \hat{y} \quad (14)$$

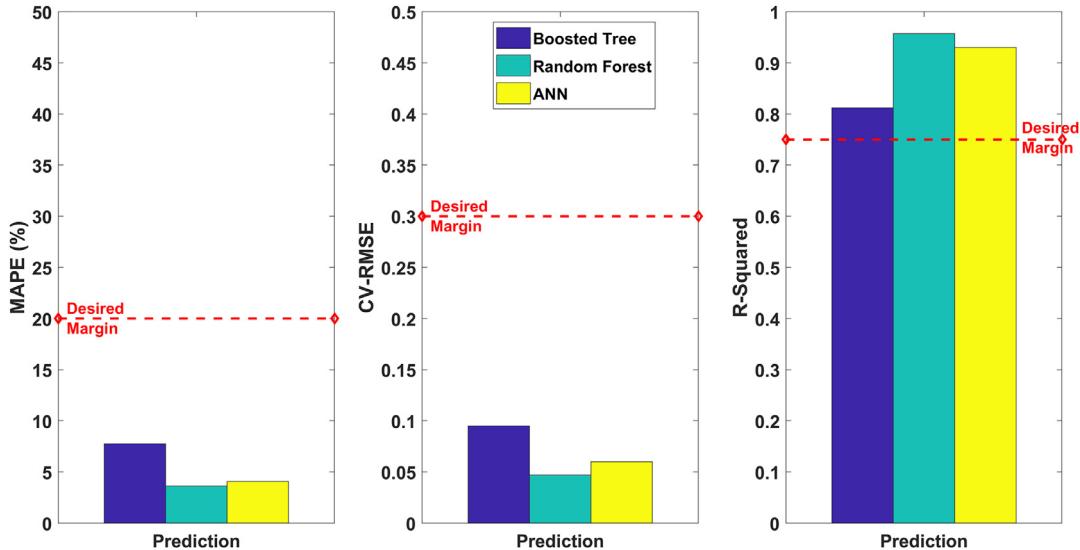


Fig. 14. Error calculation for the sum of individual predictions.

4. Conclusion

This study has focused on presenting the added value of prediction that takes into account the buildings at the group level compared to prediction at the individual level. Firstly, different ML prediction algorithms were evaluated and tested on a campus neighbourhood. From the results of the trained models, it was observed boosted-tree, and random forest outperformed the other considered regression algorithms. Considering computational power and time as well as error accuracy, boosted-tree, random forest, and ANN models were chosen for electricity demand prediction. Even though ANN did not produce exact performance in the training model when compared with the regression trees algorithms, in the prediction stage, all these three models performed similarly.

From the results of the predictions of individual buildings, it was observed that for some buildings the potential of creating a prediction model was not satisfactory. This was because of the inconsistency of the collected data during the considered period. For cases with unsatisfactory results, adequate amount of data collection for a certain period or advanced data preprocessing methods should be used. Nevertheless, most of the time building associated professionals or building owners do not acquire the required knowledge to perform such time-series analysis methods. In that case, the prediction of energy profiles in the cluster scale can help in alleviating the associated problems and lead to very high accuracy predictions. For neighborhoods such as university campuses that already acquire specific boundaries, it is beneficial to analyze its buildings at cluster level than individual scale. If the single building prediction is a prerequisite, then the unpredictable buildings can still be grouped to obtain a better forecast.

Furthermore, with the shift from the traditional grid and analog meter towards the smart grid and smart meters, the amount of data generated from buildings will increase significantly. The main question is deciding the levels at which the data should be used, the required computational speed and the required architecture for this conversion and which actors should be involved to make use of these data sets in an economical manner. In that context, the use of the prediction models on a neighborhood level to decrease the amount of time required for the computation and use of the existing computer technologies without upgrading into

advanced technologies will be beneficial for the building owners. Clustering the buildings will help in reducing the accumulation of data in significant amounts. Each building generates a lot of data, which is difficult to process without time-consuming measures. Clustering will decrease the amount of collected data, but the result for the grid side which is the predicted load will still be the same.

Another important aspect of this study is the illustration of prediction results using hourly data resolutions, instead of predicting daily or weekly. The proposed algorithms can be integrated into the building management systems for day-ahead demand and on-site renewable energy production forecasts. According to the predictions if the buildings are coordinated in advance, informed and more efficient decisions would be made in real-time to optimally utilize the on-site produced renewable electricity, and to allow the operation of the energy systems with some flexibility. Since energy neutrality is a significant concern of buildings nowadays, moving the boundaries towards neighborhood scale and making the neighborhood energy neutral will be essential and also economical for all the involved decision-maker parties.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

CRediT authorship contribution statement

Shalika Walker: Conceptualization, Methodology, Software, Validation, Writing - original draft, Writing - review & editing, Visualization. **Waqas Khan:** Software, Validation, Writing - original draft, Writing - review & editing. **Katarina Katic:** Supervision. **Wim Maassen:** Funding acquisition, Resources. **Wim Zeiler:** Funding acquisition, Supervision.

Acknowledgement

The research work is funded by NWO Perspective program TTW Project B (14180) – “Interactive energy management systems and lifecycle performance design for energy infrastructures of local communities” (<https://ses-be.tue.nl/>). We wish to express our gratitude to all the organizations which have been a part of this program. Special thanks is given to Robbert van Rossem for helping with the data collection process. Furthermore, we would like to thank dr. P.J. (Paulo) de Andrade Serra for his mathematical insights.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.enbuild.2019.109705](https://doi.org/10.1016/j.enbuild.2019.109705).

Reference

- [1] J. Kneifel, D. Webb, Predicting energy performance of a net-zero energy building: a statistical approach, *Appl. Energy.* 178 (2016) 468–483, doi:[10.1016/j.apenergy.2016.06.013](https://doi.org/10.1016/j.apenergy.2016.06.013).
- [2] S. Walker, T. Labedan, G. Boxem, W. Maassen, W. Zeiler, An assessment methodology of sustainable energy transition scenarios for realizing energy neutral neighborhoods, *Appl. Energy.* 228 (2018) 2346–2360, doi:[10.1016/j.apenergy.2018.06.149](https://doi.org/10.1016/j.apenergy.2018.06.149).
- [3] C. Robinson, B. Dilksina, J. Hubbs, W. Zhang, S. Guhathakurta, M.A. Brown, R.M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, *Appl. Energy.* 208 (2017) 889–904, doi:[10.1016/j.apenergy.2017.09.060](https://doi.org/10.1016/j.apenergy.2017.09.060).
- [4] R.G. Junker, A.G. Azar, R.A. Lopes, K.B. Lindberg, G. Reynders, R. Relan, H. Madsen, Characterizing the energy flexibility of buildings and districts, *Appl. Energy.* 225 (2018) 175–182, doi:[10.1016/j.apenergy.2018.05.037](https://doi.org/10.1016/j.apenergy.2018.05.037).
- [5] Y. Sun, G. Huang, X. Xu, A.C.-K. Lai, Building-group-level performance evaluations of net zero energy buildings with non-collaborative controls, *Appl. Energy.* 212 (2018) 565–576, doi:[10.1016/j.apenergy.2017.11.076](https://doi.org/10.1016/j.apenergy.2017.11.076).
- [6] S. Seyedzadeh, F.P. Rahimian, I. Glesk, M. Roper, Machine learning for estimation of building energy consumption and performance: a review, *Vis. Eng.* 6 (2018) 5, doi:[10.1186/s40327-018-0064-7](https://doi.org/10.1186/s40327-018-0064-7).
- [7] S.S.W. Walker, W. Lombardi, S. Lesecq, S. Roshany-Yamchi, Application of distributed model predictive approaches to temperature and CO₂ concentration control in buildings, *IFAC-PapersOnLine* (2017) 50, doi:[10.1016/j.ifacol.2017.08.107](https://doi.org/10.1016/j.ifacol.2017.08.107).
- [8] K. Katić, R. Li, J. Verhaart, W. Zeiler, Neural network based predictive control of personalized heating systems, *Energy Build.* 174 (2018) 199–213, doi:[10.1016/j.enbuild.2018.06.033](https://doi.org/10.1016/j.enbuild.2018.06.033).
- [9] M. Manfren, P. Caputo, G. Costa, Paradigm shift in urban energy systems through distributed generation: methods and models, *Appl. Energy.* 88 (2011) 1032–1048, doi:[10.1016/j.apenergy.2010.10.018](https://doi.org/10.1016/j.apenergy.2010.10.018).
- [10] Z. O'Neill, C. O'Neill, Development of a probabilistic graphical model for predicting building energy performance, *Appl. Energy.* 164 (2016) 650–658, doi:[10.1016/j.apenergy.2015.12.015](https://doi.org/10.1016/j.apenergy.2015.12.015).
- [11] C. Deb, F. Zhang, J. Yang, S.E. Lee, K.W. Shah, A review on time series forecasting techniques for building energy consumption, *Renew. Sustain. Energy Rev.* 74 (2017) 902–924, doi:[10.1016/j.rser.2017.02.085](https://doi.org/10.1016/j.rser.2017.02.085).
- [12] H. Zhao, F. Magoulès, A review on the prediction of building energy consumption, *Renew. Sustain. Energy Rev.* 16 (2012) 3586–3592, doi:[10.1016/j.rser.2012.02.049](https://doi.org/10.1016/j.rser.2012.02.049).
- [13] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time series analysis : forecasting and control*, J. Wiley & Sons (2008).
- [14] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, *Appl. Energy.* 195 (2017) 222–233, doi:[10.1016/j.apenergy.2017.03.064](https://doi.org/10.1016/j.apenergy.2017.03.064).
- [15] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, R. Saidur, A review on applications of ANN and SVM for building electrical energy consumption forecasting, *Renew. Sustain. Energy Rev.* 33 (2014) 102–109, doi:[10.1016/j.rser.2014.01.069](https://doi.org/10.1016/j.rser.2014.01.069).
- [16] E. Mocanu, P.H. Nguyen, M. Gibescu, W.L. Kling, Deep learning for estimating building energy consumption, *Sustain. Energy, Grids Networks.* 6 (2016) 91–99, doi:[10.1016/j.segan.2016.02.005](https://doi.org/10.1016/j.segan.2016.02.005).
- [17] Y. Huang, Y. Yuan, H. Chen, J. Wang, Y. Guo, T. Ahmad, A novel energy demand prediction strategy for residential buildings based on ensemble learning, *Energy Procedia* 158 (2019) 3411–3416, doi:[10.1016/j.egypro.2019.01.935](https://doi.org/10.1016/j.egypro.2019.01.935).
- [18] Z. Wang, Y. Wang, R.S. Srinivasan, A novel ensemble learning approach to support building energy use prediction, *Energy Build.* 159 (2018) 109–122, doi:[10.1016/j.enbuild.2017.10.085](https://doi.org/10.1016/j.enbuild.2017.10.085).
- [19] E. Mocanu, P.H. Nguyen, W.L. Kling, M. Gibescu, Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning, *Energy Build.* 116 (2016) 646–655, doi:[10.1016/j.enbuild.2016.01.030](https://doi.org/10.1016/j.enbuild.2016.01.030).
- [20] X. Xu, W. Wang, T. Hong, J. Chen, Incorporating machine learning with building network analysis to predict multi-building energy use, *Energy Build.* 186 (2019) 80–97, doi:[10.1016/j.enbuild.2019.01.002](https://doi.org/10.1016/j.enbuild.2019.01.002).
- [21] F. Smarra, A. Jain, T. de Rubeis, D. Ambrosini, A. D’Innocenzo, R. Mangharam, Data-driven model predictive control using random forests for building energy optimization and climate control, *Appl. Energy.* 226 (2018) 1252–1272, doi:[10.1016/j.apenergy.2018.02.126](https://doi.org/10.1016/j.apenergy.2018.02.126).
- [22] C.E. Kontokosta, C. Tull, A data-driven predictive model of city-scale energy use in buildings, *Appl. Energy.* 197 (2017) 303–317, doi:[10.1016/j.apenergy.2017.04.005](https://doi.org/10.1016/j.apenergy.2017.04.005).
- [23] A.B.M.S. Ali, S. Azad, in: *Demand Forecasting in Smart Grid*, Springer, London, 2013, pp. 135–150, doi:[10.1007/978-1-4471-5210-1_6](https://doi.org/10.1007/978-1-4471-5210-1_6).
- [24] S.S. Ahmed, R. Thiruvengadam, A.S. Shashank Karrthikeyaa, V. Vijayaraghavan, in: *A Two-Fold Machine Learning Approach for Efficient Day-Ahead Load Prediction at Hourly Granularity for NYC*, Springer, 2020, pp. 84–97, doi:[10.1007/978-3-030-12385-7_8](https://doi.org/10.1007/978-3-030-12385-7_8).
- [25] N.G. Paterakis, E. Mocanu, M. Gibescu, B. Stappers, W. van Alst, Deep learning versus traditional machine learning methods for aggregated energy demand prediction, in: *2017 IEEE PES Innov. Smart Grid Technol. Conf. Eur.*, IEEE, 2017, pp. 1–6, doi:[10.1109/ISGTEurope.2017.8260289](https://doi.org/10.1109/ISGTEurope.2017.8260289).
- [26] Statistics and Machine Learning Toolbox Documentation, MathWorks BENELUX. (n.d.). <https://nl.mathworks.com/help/stats/> (accessed October 18, 2019).
- [27] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, *Energy* 32 (2007) 1761–1768, doi:[10.1016/j.energy.2006.11.010](https://doi.org/10.1016/j.energy.2006.11.010).
- [28] M.W. Ahmad, M. Moursched, Y. Rezgui, Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression, *Energy* 164 (2018) 465–474, doi:[10.1016/j.energy.2018.08.207](https://doi.org/10.1016/j.energy.2018.08.207).
- [29] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random forest based hourly building energy prediction, *Energy Build.* 171 (2018) 11–25, doi:[10.1016/j.enbuild.2018.04.008](https://doi.org/10.1016/j.enbuild.2018.04.008).
- [30] H. Taghavifar, A. Mardani, A comparative trend in forecasting ability of artificial neural networks and regressive support vector machine methodologies for energy dissipation modeling of off-road vehicles, *Energy* 66 (2014) 569–576, doi:[10.1016/j.energy.2014.01.022](https://doi.org/10.1016/j.energy.2014.01.022).
- [31] A. Kialashaki, J.R. Reisel, Development and validation of artificial neural network models of the energy demand in the industrial sector of the United States, *Energy* 76 (2014) 749–760, doi:[10.1016/j.energy.2014.08.072](https://doi.org/10.1016/j.energy.2014.08.072).
- [32] F. Ascione, N. Bianco, C. De Stasio, G.M. Mauro, G.P. Vanoli, Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: a novel approach, *Energy* 118 (2017) 999–1017, doi:[10.1016/j.energy.2016.10.126](https://doi.org/10.1016/j.energy.2016.10.126).
- [33] M.A.R. Biswas, M.D. Robinson, N. Fumo, Prediction of residential building energy consumption: a neural network approach, *Energy* 117 (2016) 84–92, doi:[10.1016/j.energy.2016.10.066](https://doi.org/10.1016/j.energy.2016.10.066).
- [34] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning*, Springer New York, 2013, doi:[10.1007/978-1-4614-7138-7](https://doi.org/10.1007/978-1-4614-7138-7).
- [35] M. Sugiyama, *Introduction to Statistical Machine Learning*, Elsevier, 2016, doi:[10.1016/C2014-0-01992-2](https://doi.org/10.1016/C2014-0-01992-2).
- [36] K. Kevin N. Gurney, Kevin, *An Introduction to Neural Networks*, UCL Press, 1997. <https://dl.acm.org/citation.cfm?id=523781> (accessed September 30, 2019).
- [37] C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: Comparison of trends in practice and research for deep learning, (2018). <https://arxiv.org/abs/1811.03378> (accessed October 18, 2019).
- [38] F. Farhadi, Learning Activation Functions in Deep Neural Networks, (2017). https://publications.polymtl.ca/2945/1/2017_FarnoushFarhadi.pdf (accessed October 18, 2019).
- [39] S. Sapna, Backpropagation Learning Algorithm based on Levenberg Marquardt Algorithm, in: 2012. doi:[10.5121/csit.2012.2438](https://doi.org/10.5121/csit.2012.2438).
- [40] V. Vacic, Summary of the training functions in Matlab’s NN toolbox, (n.d.). <https://docplayer.net/42096520-Summary-of-the-training-functions-in-matlab-s-nn-toolbox-vladimir-vacic.html> (accessed October 18, 2019).
- [41] Y.-J. Ma, M.-Y. Zhai, Day-Ahead prediction of microgrid electricity demand using a hybrid artificial intelligence model, *Processes* 7 (2019) 320, doi:[10.3390/pr7060320](https://doi.org/10.3390/pr7060320).
- [42] D. Benouada, F. Murtagh, J.L. Starck, O. Renaud, Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting, *Neurocomputing* 70 (2006) 139–154, doi:[10.1016/j.neucom.2006.04.005](https://doi.org/10.1016/j.neucom.2006.04.005).
- [43] J. Runge, R. Zmeureanu, Forecasting energy use in buildings using artificial neural networks: a review, *Energies* 12 (2019) 1–27.
- [44] R. Sevlian, R. Rajagopal, A scaling law for short term load forecasting on varying levels of aggregation, *Int. J. Electr. Power Energy Syst.* 98 (2018) 350–361, doi:[10.1016/j.ijepes.2017.10.032](https://doi.org/10.1016/j.ijepes.2017.10.032).
- [45] C. Kasemset, N. Sae-Haew, A. Sopadang, Multiple regression model for forecasting quantity of supply of off-season longan, *C. J. Nat. Sci.* 13 (2014) 391, doi:[10.12982/cmjns.2014.0044](https://doi.org/10.12982/cmjns.2014.0044).
- [46] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, *Renew. Sustain. Energy Rev.* 73 (2017) 1104–1122, doi:[10.1016/j.rser.2017.02.023](https://doi.org/10.1016/j.rser.2017.02.023).
- [47] A. Taebi, H. Mansy, A. Taebi, H.A. Mansy, Time-Frequency distribution of seismocardiographic signals: a comparative study, *Bioengineering* 4 (2017) 32, doi:[10.3390/bioengineering4020032](https://doi.org/10.3390/bioengineering4020032).

- [48] C. Fan, Y. Ding, Cooling load prediction and optimal operation of HVAC systems using a multiple nonlinear regression model, Energy Build 197 (2019) 7–17, doi:[10.1016/J.ENBUILD.2019.05.043](https://doi.org/10.1016/J.ENBUILD.2019.05.043).
- [49] A. Guideline, Measurement of energy and demand savings, 2002. http://www.eeperformance.org/uploads/8/6/5/0/8650231/ashrae_guideline_14-2002_measurement_of_energy_and_demand_saving.pdf (accessed March 27, 2019).
- [50] I.S. Sacramento, G.P. Souza, R.S. Wazlawick, Principal component analysis to reduce forecasting error of industrial energy consumption in models based on neural networks, in: Springer LNAI 8467 - Artif. Intell. Soft Comput., Springer, Cham, 2014, pp. 143–154, doi:[10.1007/978-3-319-07173-2_14](https://doi.org/10.1007/978-3-319-07173-2_14).
- [51] M. Feurer, F. Hutter, Hyperparameter optimization, Autom. Mach. Learn (2019) 3–33, doi:[10.1007/978-3-030-05318-5_1](https://doi.org/10.1007/978-3-030-05318-5_1).
- [52] D.L.J. Alexander, A. Tropsha, D.A. Winkler, Beware of r^2 : simple, unambiguous assessment of the prediction accuracy of qsar and qspr models, J. Chem. Inf. Model. 55 (2015) 1316–1322, doi:[10.1021/acs.jcim.5b00206](https://doi.org/10.1021/acs.jcim.5b00206).