Vol. 34, No. 3 (2024) | DOI: 10.37190/ord240309

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

Analysing and forecasting the energy consumption of healthcare facilities in the short and medium term. A case study

Ali Koç[1*] Serap Ulusam Seçkiner[1]

[1] *Department of Industrial Engineering, Gaziantep University, Gaziantep, Turkey*

[*] *Corresponding author, email address: Ali.Koc1@saglik.gov.tr*

Abstract

Healthcare facilities consist of multiple large buildings with complex energy systems and high energy consumption, resulting in high carbon emissions. The increasing trend in energy consumption of these facilities and the process of selecting an energy supplier from the open market requires reliable and robust energy forecasting studies. This situation calls for the use of reliable and accurate energy consumption prediction models for the energy needs of healthcare buildings. The aim of this study is to present a prediction framework based on historical energy consumption at different time intervals using six supervised regression algorithms, three linear single, one non-linear single and two non-linear ensembles. The approach adopted for predicting hospital energy consumption involves five steps: data acquisition, data pre-processing, data prediction, hyper-parameter optimisation and feature analysis. Furthermore, all regression algorithms have undergone hyper-parameter optimisation using random search, grid search and Bayesian optimisation to achieve the minimum prediction errors represented by different metrics. The results displayed that the two ensemble models, Extreme Gradient Boosting and Random Forest, outperformed single models in hourly, daily, and monthly energy load prediction. Nevertheless, when considering the computational time for all regression models, the single models have better computational times, although the error metrics are not as good as for the ensemble models. In addition, grid search and Bayesian optimisation performed better than random search in finding optimal hyperparameter values for all datasets. Finally, thanks to feature importance analysis, the most influential features under the hourly, daily, and monthly electrical and monthly natural gas prediction were identified.

Keywords: *healthcare facilities, electricity, natural gas, consumption, forecasting, machine learning*

1. Introduction

Factors such as limited resources, excessive and growing consumption, and high costs make it imperative to adopt an energy management perspective and use energy efficiently. In this context, large buildings represent some of the great consumers of energy. Hospitals, hotels, large sports centres and other buildings in particular are major energy consumers, responsible for 40% of energy consumption and 36% of

Received 4 August 2023, accepted 14 July 2024 , published online 17 October 2024
ISSN 2391-6060 (Online)/© 2024 Authors

The costs of publishing this issue have been co-financed by the program *Development of Academic Journals* of the Polish Ministry of Education and Science under agreement RCN/SP/0241/2021/1

CO_2 emissions in the EU [16]. The increasing trend in the energy demand of such buildings requires accurate and robust energy consumption forecasts, which should contribute to effective planning, long-term strategies and control of energy consumption in the building sector. According to Eckelman and Sherman [15], healthcare facilities are the second largest energy-consuming commercial buildings in the US, after food service. In such energy-intensive buildings like healthcare facilities, the benefits of accurate energy demand forecasting are that it helps healthcare facility managers to make reliable energy budget projections and to select appropriate suppliers from the open market.

This study aims to analyse the dynamics and predictability of electricity and natural gas consumption in healthcare buildings, where energy costs and demand are mostly subject to uncertainty. In addition to uncertainty, the drawbacks associated with the lack of data make it difficult to implement forecasting techniques for predicting energy loads in healthcare facilities. As machine learning regression algorithms are trained on accurate data from past periods and future predictions are made by finding appropriate parameters, data accuracy is a very important factor in the application of these algorithms. Furthermore, when electricity or natural gas is supplied from the open market based on forecasts for a hospital facility, where continuous energy and electricity consumption is required due to the continuous use of technological equipment, the accuracy of the ML algorithm forecasts becomes more and more important.

In the context of the difficulties mentioned above, in this study, electricity and natural gas demand forecasting procedures have been presented based on different datasets, namely hourly, daily, and monthly time intervals, since it is desired to observe the impact of time granularity on the forecasting performance. For each time granularity, different inputs were considered based on their impact on energy consumption. For short-term time granularities, we have mainly considered weather-related inputs and variables such as intraday period and day type, while for medium-term time granularities, input variables such as number of patients, bed occupancy rate and unit price have been considered with a broader perspective.

In our study, forecasting models for electricity and natural gas consumption in healthcare buildings have been developed based on linear single, non-linear single, and non-linear ensemble supervised machine learning algorithms by simultaneous considering multi-factors as mentioned above. The performance of the prediction algorithms was evaluated using statistical metrics such as root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and R2. In addition, hyper-parameter optimization was carried out using grid search and genetic algorithms to achieve a minimum performance error value with the regression algorithms.

The rest of the paper is organized as follows. Section 2 is about a comprehensive literature and its review. Section 3 presents the data and methodology. Section 4 presents the results and discussion of a case study application based on hospital data. The final section 5 presents the conclusion and future work.

2. Literature review

In the context of the increasing trend of energy demand in buildings, many studies have been carried out in terms of prediction of building energy consumptions such as electricity natural gas, heating, and cooling by using different forecasting methods. There are many studies in the literature on energy demand forecasting for different building types, such as residential [5, 41], university academic buildings [36, 45], office buildings [1, 29, 46], commercial buildings [26, 37, 44], and healthcare buildings [3, 8, 20, 31, 40]. These studies considered different time horizons for forecasting, which can be represented as very short-term, short-term, and medium-term. The very short term can be categorized as a time granularity of less than 1 hour [3, 20, 29, 37]. The short term can be considered as hourly, daily, and weekly time granularities [2, 11, 31, 33], while the middle term can be considered as a time horizon between one month and one year [4, 40, 41].

Table 1. Related studies

Reference	Technique	Ref. data set	Energy type	Sampling interval
[33]	ANN and EenrgyPlus	university building	electricity–heating	daily
[31]	ANN	healthcare facility	electricity	hourly
[27]	SVR and modified firefly aAlgorithm	cities	electricity load	daily
[3]	ANN	healthcare facility	electricity	minute–hourly–daily
[5]	ANN	houses	electricity	daily
[46]	wavelet transform, SVR, PLS regression	office	heating–cooling	hourly–daily
[20]	PCA analysis, autoregressive, OPLS	healthcare facility	electricity	minute
[2]	MR, GP, ANN, DNN and SVR	university building	electricity	daily
[44]	ANN, Fourier's law based analysis, TRNSYS	commercial Ckyscraper	cooling	hourly
[45]	ANN	university building	electricity	hourly
[1]	tree bagger, GPR, MLR, ANN	office buildings	heating–cooling	hourly–daily–monthly
[26]	ANN	comnercial building	electricity–natural gas–cooling	hourly
[11]	MLR and TRNSYS	commercial building	heating–cooling	hourly
[4]	MLR, ANN and SVR	City	Natural Gas	monthly
[13]	ANN, SVR, gradient boosting	healthcare facility	cooling	hourly
[21]	ANN	country	electricity	minute
[36]	LSTM, sinecosine algorithm	university building	electricity	hourly–monthly–yearly
[40]	MLR, ANN, SVR	healthcare facility	electricity	monthly
[12]	least squares, Cobb–Douglas	country	natural gas	yearly
[39]	multi-task learning, LSSVR	industrial building	electricity, heating, cooling and natural gas	hourly
[8]	MLR, ridge, lasso, ElasticNet SVR, Gaussian, Random Forest, XGBoost	healthcare facility	electricity	daily–weekly
[25]	ARIMA, ANN, ELM, TBATS	cities	natural gas	daily–weekly
[22]	ANN, SVR, SARIMAX, DNN, LSTM	commercial building	electricity	daily–monthly
[14]	MLR, Buckingham theorem, ANN	commercial building	heating	hourly
[41]	LSSVR, RBFNN, symbiotic organism search	residential building	electricity, heating, cooling	daily–monthly
[29]	A3C, DDPG, RDPG	office buildings	electricity load	minute
[37]	k-mean clustering, LSTM	commercial building	electricity	minute
[28]	grey theory, GM(1, 1), TBGM(1, 1)	city	electricity	yearly
[30]	adaptive LSTM, genetic algorithm	university building	electricity	hourly
[19]	Cox proportional hazards model	healthcare facility	electricity–heating	monthly
[32]	Hybrid simulation approach (EnergyPlus + ANN + GM)	healthcare facility	cooling	daily
[38]	WOA-BiLSTM model	healthcare facility	electricity	hourly

In addition to studies on building energy prediction, many studies have been conducted for electricity demand forecasting [21, 27, 28] and natural gas demand forecasting [4, 12, 25] of countries or cities with datasets based on different time granularities. For example, Tatoglu et al. attempted to predict the monthly natural gas consumption of Istanbul city using machine-learning methods such as multiple linear regression, artificial neural networks, and Support Vector Regression [4]. They find that the best forecasting method for natural gas consumption is Support Vector Regression with a MAPE value of 5.53%. As shown in Table 1, there is a wide range of forecasting approaches used for energy in the literature, including engineering, statistical,and machine learning methods.

Of the machine learning methods used, as shown in Table 1, ANN [13, 21, 31, 33, 34] and its variants [30, 37] are the most widely used machine learning algorithm for energy forecasting because they are flexible, non-linear, and applicable with different activation functions under different hidden layers. In recent years, many studies have tried to find the algorithm with the lowest prediction error by trying more than one algorithm, rather than using only one algorithm [8, 22, 29, 40].

Healthcare facilities have quite complex buildings with huge electricity, heating, and cooling consumption like an industrial factory. Although energy forecasting studies for such buildings are scarce in the literature, a number of them display an increasing trend. Zorita-Lamadrid et al. [31] tried to predict hourly electricity load using an ANN-Multilayer Perceptron (MLP) approach, similarly Silvestro et al. [3] also developed an ANN-MLP model based on a backpropagation algorithm to predict the very short term (minute) electricity load of a large hospital facility. Gordillo-Orquera et al. [20] used different statistical models such as Principal Component Analysis (PCA) and Autoregressive (AR), Orthonormal Partial Least Squares (OPLS) for short-term electricity load forecasting of a hospital and a primary care center, respectively. Dulce et al. [13] focused on predicting of the thermal cooling demand of a hospital adopting a genetic methodology that searches for low complexity models through feature selection, parameter tuning, and parsimonious model selection. They tested their methodology using artificial neural networks, Support Vector Regression, and gradient boosting techniques. Zor et al. [40] used multiple linear regression (MLR), Artificial Neural Network (ANN), and Support Vector Regression (SVR) techniques to predict the long-term electricity consumption of a hospital, while Zhang et al. [8] took into consideration eight different regression algorithms including MLR, Random Forest to project daily and weekly electricity load consumption of a hospital.

As for summation of the literature, it has generally been observed that while the advanced deep learning and machine learning techniques have been preferred for forecasting problems, the large number of parameters for these techniques give rise to complexity in application. On the other hand, time series and regression models, which are easier to apply in practice, have been found to be highly preferred for forecasting problems. In terms of temporal granularity, studies dealing with short-term forecasting problems have generally used deep learning or machine learning techniques, while long-term studies have come forward regression models or time series models (especially those related to ARIMA).

3. Materials and methods

The research methodology involves the execution of different machine learning algorithms, which have been applied to stand-alone datasets with different temporal granularity, hourly, daily and monthly for electricity and monthly for natural gas, respectively, as shown in Figure 1. The methodology consists of five main steps: (1) data collection, data preparation, splitting the data into training and test sets, (2) training ML regression algorithms with the prepared data, (3) optimization of hyperparameters for all models,

(4) analysis of prediction and computation time performance, and (5) discussion of the importance of each feature.

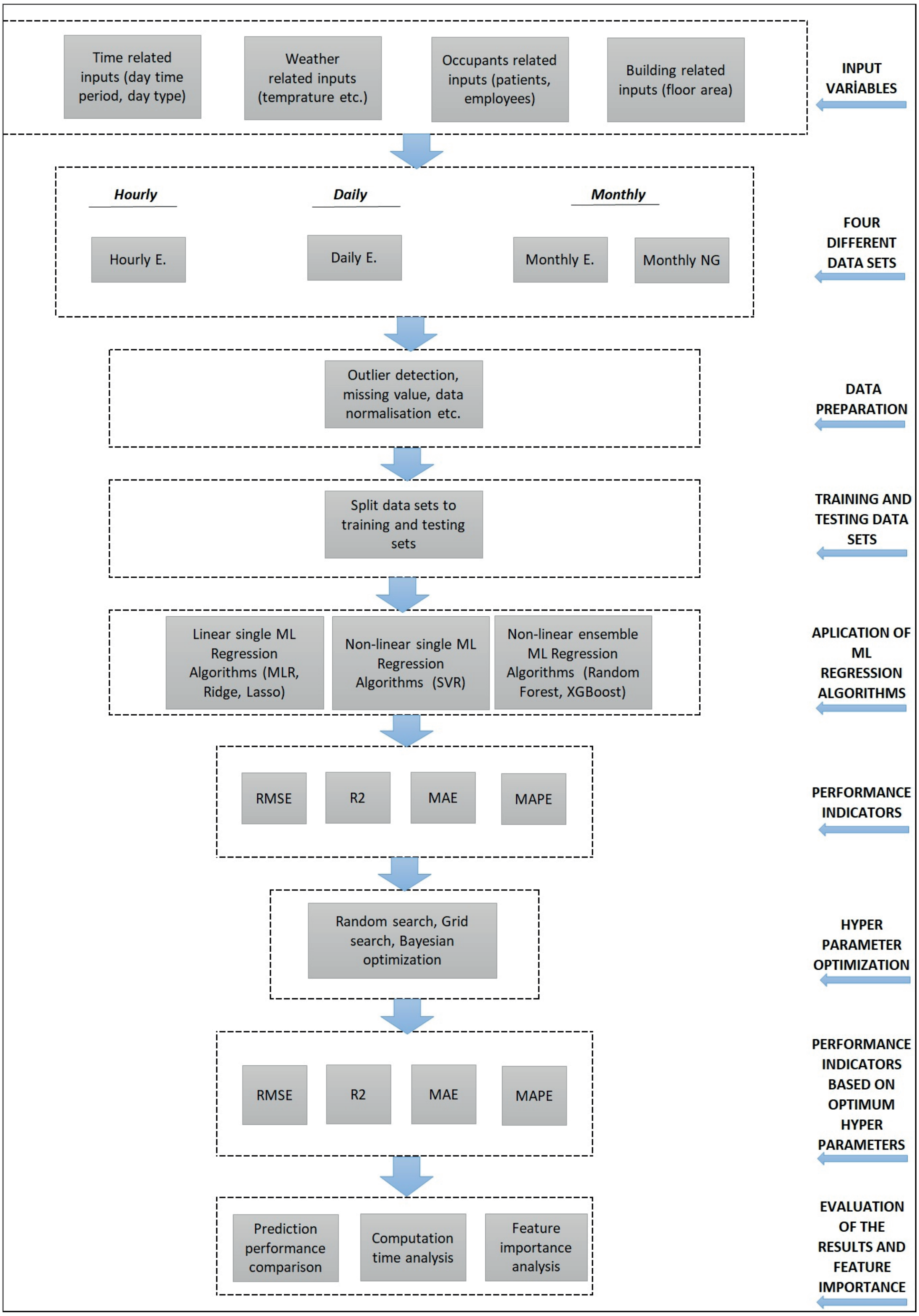

Figure 1. Flowchart of energy consumption forecasting of the hospital

In this study, two main types of energy, electricity and natural gas, have been treated separately as they are the source of different loads such as electricity, heating and cooling. Three temporal granularities were used for electricity: hourly, daily and monthly, while for natural gas only monthly data were used due to a lack of detailed data. The input variables used for each dataset are shown in Table 2. The dataset of each module was divided into training and test sets as shown in Table 3. For all datasets, the training and testing datasets have been generated in a time-series manner to correctly reflect the seasonal effect, as 70% for training and 30% for testing. The generated training dataset is then used to train the machine learning regression models for each module, and the testing dataset is used to evaluate the performance of the trained models.

Table 2. The inputs and output for forecasting models

Input variable	Hourly-E (output)	Daily-E (output)	Monthly-E (output)	Monthly-NG (output)
Day time period (day, night, peak)	×			
Outdoor air dry-bulb temperature, °C	×	×		
Outdoor air feel-like temperature, °C	×	×		
Relative humidity, %	×			
Wind speed, km/h	×			
Day type (working, holiday)	×	×		
Heating degree days (HDD), (severity of the cold)	×			×
Cooling degree days (CDD), (need for cooling)		×	×	
Unit price, $/kWh			×	×
Number of inpatients			×	×
Number of outpatients			×	×
Bed occupancy rate			×	×
Number of employees			×	×
Floor area, m^2			×	×
Season type			×	×

Table 3. Training and testing samples

Dataset	Training period	Testing dataset
Monthly natural gas	January 2012–December 2016	January 2017–December 2018
Monthly electricity	January 2012–December 2016	January 2017–December 2018
Daily electricity	1st January 2018–4th September 2018	5th September 2016–31st December 2018
Hourly electricity	first 6132 data point	last 2628 data point

Healthcare facilities have buildings that use all four forms of energy, electricity, heating, hot water and cooling, together and intensively. In hospitals, the increasing energy consumption as a function of the density of use and the size of the structure is very important in terms of both cost and efficient use of resources. Furthermore, accurate estimation of energy loads or consumption has a serious impact on the efficient use of energy. In this context, the forecasting methods were applied to predict the energy consumption of our reference hospital, which is a general hospital located in the Eastern Anatolia region of Turkey.

3.1. Data collection and input variables

This study focused mainly on three time periods in terms of forecasting horizon, based on hourly, daily and monthly periods. In order to analyze energy consumption from a compact perspective, we have treated both electricity and natural gas individually. Each temporal granularity is influenced by different variables that have an impact on the target output variable. In this context, we have considered three different time periods for electricity forecasting models: hourly, daily and monthly. The hourly electricity module has been considered with the aim of observing the effect of time-of-use (TOU) on electricity consumption, which consists of three different time periods day, night and peak. Based on the Turkish electricity market and its regulations, (1) the night period includes the hours 22:00–06:00, (2) the day period includes the hours 06:00–17:00, while (3) the peak period includes the hours 17:00–22:00. In addition, the process of selecting a supplier of electricity from the open market requires time-of-use based consumption in order to consider a choice of single or multiple time tariff options. We have also considered daily and monthly electricity consumption. For natural gas, on the other hand, we have only used monthly time periods due to a lack of detailed data based on other time granularities.

The monthly data used in this study for both electricity and natural gas were collected from the hospital's electricity and natural gas bills, which cover the period 2012-2018. On the other hand, the hourly and daily electricity data were collected based on a combination of information from different sources. In fact, hourly and daily energy consumption wasn't recorded in our reference hospital due to the lack of intelligent building energy systems. To overcome this obstacle, especially for electricity, we used hourly electricity data from previous years obtained from an electricity company as part of a feasibility study carried out in the reference hospital. These data were adapted to the year 2018, with hourly and daily coefficients that take into account whether the day is a working day or a holiday. In this context, this study used 2018 annual data for electricity in terms of hourly and daily forecasts. As shown in Table 3, the different input variables are used for each time horizon. In general, energy consumption in buildings is influenced by four main combinations of inputs: weather-related inputs, occupant-related inputs, building-related inputs and time-related inputs. For hourly electricity consumption, we focused on weather-related inputs and time-related inputs. Weather-related explanatory variables include outdoor dry bulb temperature, outdoor feel-good temperature, relative humidity and wind speed, while time-related input variables are categorical variables such as time of day and type of day. As mentioned above, time-of-use has a major impact on electricity consumption. We have also taken into account working days and holidays, which have a significant impact on energy consumption. On the other hand, daily electricity consumption is influenced by weather variables such as average daily outdoor temperatures and cooling degree days. In addition, the type of day was treated as a time-related input for the daily forecast.

Monthly datasets have been produced for both electricity and natural gas. For the monthly consumption models, we prioritized occupancy-related inputs such as inpatients, outpatients, bed occupancy and number of employees. In addition, floor area, degree days and monthly unit energy prices were also considered as inputs for the monthly models.

Due to the complexity of hospital energy infrastructure, the choice of explanatory variables is a very important step in modelling forecasting methods. In regression models, some variables are considered as explanatory variables and the target variable is considered as the dependent variable. An explanatory variable must have an impact on the target value at a meaningful statistical level. To explore this effect, we used Pearson correlation analysis, which ensures that it is possible to decide whether input extraction from the model is necessary. The effect of the input variables shown in Table 3 and mentioned above will be analyzed in the next section.

3.2. Data preparation

The data preparation section mainly includes various pre-processes such as outlier detection, missing value extraction, correlation/feature extraction and data normalisation. Data pre-processing was performed to prepare the input and output datasets for more accurate prediction models.

3.2.1. Outlier detection

An outlier is an observation that appears distant and deviates from the overall pattern in a sample. The outlier detection method is considered to eliminate the potential outliers in the raw data to obtain qualified data. There are various approaches to outlier detection and in this context we have adopted the modified Z-score method [23] to eliminate outlier values from datasets. In this method, the median and the median of the absolute deviation of the median (MAD) are used instead of the mean and the standard deviation of the sample. For the dataset of the size m

$$|x_t - \tilde{x}| \ \forall_t = 1, 2_{\text{ani}}\, m \tag{1}$$

$$\text{MAD} = \text{median}\left\{|x_t - \tilde{x}|\right\} \tag{2}$$

$$\text{modified } Z - \text{ score } = M_i = \frac{0.6745\,(x_t - \tilde{x})}{\text{MAD}} \tag{3}$$

$$x_t = \begin{cases} \text{mean}(X) + 2\,\text{std}(X) & \text{if } M_i > 3.5 \text{ (an outlier)} \\ x_t & \text{otherwise} \end{cases} \tag{4}$$

where $\tilde{x}$ is the median of sample dataset, X is the random variable of all x_t values, $\text{mean}(X)$ is the average value of X, whereas $\text{std}(X)$ is the standard deviation of X. Equation (1) is a new data set obtained from an absolute difference between each value and the median. Equation (2) is the new median of the new data set. The modified Z-score is acquired by equation (3). After outliers are determined, there are a few options related to evaluating the outliers, where one of them is an assignment of a new value to the outlier. In this context, we have adopted equation (4). Outlier analysis has been conducted for historical meteorological data in particular.

3.2.2. Missing values

The missing values in our datasets are added to dataset using a data interpolation method as shown below. Equation (5) ensures to find missing values with arithmetic mean of two consecutive available values.

$$x_t = \begin{cases} \dfrac{x_{t-1} + x_{t+1}}{2} & x_t \text{ is missing value, } x_{t-1} \text{ and } x_{t+1} \text{ are available values} \\ \dfrac{x_{t-1} + x_{t-2}}{2} & x_t \text{ and } x_{t+1} \text{ are missing values} \end{cases} \tag{5}$$

3.2.3. Correlation analysis and feature extraction

Correlation analysis is an approach to confirming the degree of relationship between two or more variables. In forecasting models, the degree of relationship between explanatory variables and target variables is significant from the point of view of model accuracy. If there is no relationship between two variables, it means that this explanatory variable does not affect the target variables. Consequently, the related input variables or characteristics must be removed from the data set. In this study, we used Pearson's correlation, which is commonly used in data analysis. Correlation values vary between -1 and +1, indicating strong negative and positive correlations, while 0 indicates no linear correlation. The correlations obtained using Pearson's analysis between the explanatory variables and the target values are shown in Table 4.

Table 4. Correlation values between energy consumption data and related inputs

Hourly electricity (2018)	Dry-bulb temp. 0.616	Feel-like temp. 0.585	Relative humidity –0.537	WindSpeed 0.165
Daily electricity (2018)	Dry-bulb temp. 0.626	Feel-like temp. 0.601	CDD 0.753	
Monthly electricity (2012–2018)	Outpatients 0.162 Elec. unit price –0.223	Inpatients 0.121 CDD 0.82	Employee 0.05 Floor area 0.226	Bed occupancy –0.331
Monthly electricity (2012–2018)	Outpatients 0.039 Nat. gas unit price 0.028	Inpatients 0.112 HDD 0.979	Employee 0.147 Floor area –.0118	Bed occupancy 0.693

Different criteria can be applied to select input variables based on the correlation results shown in Table 4. In this study, the frontier correlation value for the selection has been accepted as 0.1. In this manner, inputs with correlation values less than 0.1 were excluded from the model. In hourly and daily datasets, the correlations between explanatory variables and target values are higher than 0.1. On the other hand, in monthly electricity data, the employee variable has been excluded from the model with its 0.050 correlation coefficient. Also in monthly natural gas data, the outpatients variable, natural gas unit price variable, and floor area variable have been excluded from the model with their low correlation coefficients.

3.2.4. Data normalization

Before the application of machine learning algorithms, data have been normalized with the min-max normalization method. Normalization is a process of making model data in a standard format so that the training is improved, accurate, and faster.

$$y_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{6}$$

where y_i is the normalized value, x_i is the sample value, $\max(x)$ is the maximum value, and $\min(x)$ is the minimum value in equation (6).

3.2.5. Seasonality adjustment for monthly datasets

A dummy variable is a numerical variable that represents categorical data in regression models. Once a categorical variable has been recoded as a dummy variable, the dummy variable can be used in regression analysis just like any other quantitative variable. This method transforms the categorical variable into a set of binary variables (also known as dummy variables). Dummy encoding uses $N-1$ features to represent N labels/categories. It is used to transform categorical variables such as type of day, time of day and type of season.

3.2.6. Dummy variables for categorical features

Seasonality is the presence of variations that occur at specific regular intervals of less than one year, such as weekly, monthly or quarterly. In this study, instead of using seasonal indices to eliminate seasonality, we have preferred to group months with similar consumption behaviour to create seasons and use them as categorical variables in the model. As shown in Figure 2 and Table 5, we have grouped the months as a season, which includes consecutive months with similar consumption that provide both electricity and natural gas consumption curves.

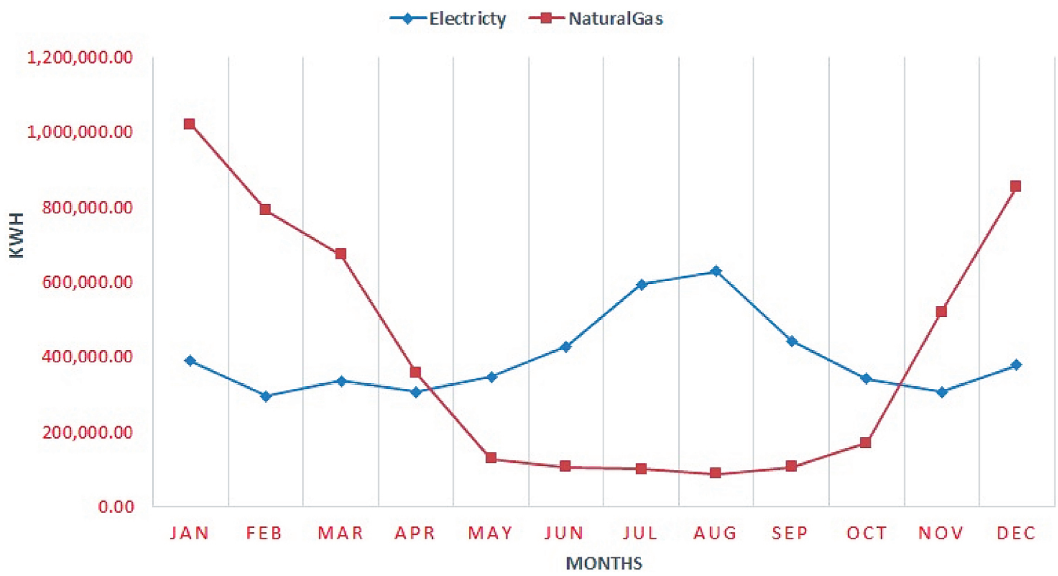

Figure 2. Monthly average energy fluctuations in consumption per one year for the hospital

Table 5. Season groups and included months

Season group	Month
Winter	January, February
Spring	March, April
May	May
Summer	June, July, August, September
October	October
Autumun	November, December

3.3. Regression algorithms in machine learning

In machine learning approaches, it is used various kinds of algorithms allow machines to learn the relationships within the data handled and make predictions based on patterns or rules identified from the dataset. Three main types of machine learning are supervised learning (predictive), unsupervised learning (descriptive), and reinforcement learning respectively [17]. Supervised learning algorithms are trained on well-labelled data and they predict the outputs based on this data. This training dataset consists of inputs and correct outputs that allow the model to learn over time. Supervised learning can be used for two types of problems which are regression and classification. In this study, we have used regression algorithms, which are used to understand the relationship between dependent and independent variables in order to make forward projections. In the other words, the future values are predicted with the help of regression algorithms in Machine Learning.

Regression algorithms can be divided into three main categories that are linear single, non-linear single, and non-linear ensemble algorithms. The ML regression algorithms used are the following.

3.3.1. Multiple linear regression

Multiple linear regression is an extension of simple linear regression because more than one explanatory variable is needed to predict the target variable. Multiple linear regression is a machine learning algorithm that allows us to examine how multiple independent (predicted) variables relate to a dependent variable. A dependent variable is modelled as a function of several independent variables with corresponding coefficients, with the constant term. The multiple regression equation can be written as follows;

$$Y = b_0 + b_1 X_1 + \cdots + b_k x_k + \varepsilon \tag{7}$$

where b_0–b_k are the regression coefficients to be estimated according to observations in equation (7). The last term in the formulation,ε, shows the random error and is referred as the residual for checking the overall significance of the model and each regression coefficient [6]. In the multiple regression analysis, we have to avoid multicollinearity between the independent variables. On the other hand, it is available some other assumptions such as normality, linearity.

3.3.2. Ridge regression

Ridge regression, a variant of linea one, is a shrinkage method commonly used in machine learning [24] to deal with the estimation of models with multicollinearity between explanatory variables. Conventional ordinary least squares (OLS) predictor aims to minimize the residual sum of squares, whereas ridge regression aims to minimize the residual sum of squares (RSS) by imposing a penalty on the OLS parameters. Mathematical representation of the ridge regression as follows [37];

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = \text{RSS} + \lambda\sum_{j=1}^{p}\beta_j^2 \tag{8}$$

where $\lambda\sum_{j=1}^{p}\beta_j^2$ is a shrinkage penalty function and λ is a tuning parameter that shrinks parameter estimates towards zero.

3.3.3. Lasso regression

Least absolute shrinkage and selection operator (LASSO) is another linear shrinkage estimator like ridge regression. It also adds a penalty for non-zero coefficients, but unlike ridge regression, which penalizes sum of squared coefficients, lasso penalizes the sum of their absolute values. equation (9) shows a general formula for lasso regression [24]

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j|=\text{RSS}+\lambda\sum_{j=1}^{p}|\beta_j| \tag{9}$$

3.3.4. Support Vector Regression

Another machine learning method support vector machine (SVM) is a data mining method put forward by Vapnik [42, 43], which is used for classification, regression, estimation, pattern recognition etc. making with linear or non-linear (with kernel functions) algorithms. In this study, we have an interest in the non-linear SVR algorithm, where the kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation. The main ideas behind the SVR are minimizing error are individualizing the hyperplane which maximizes the margin [35].

For example, we have a training data $\{(x_1, x_1), \ldots, (x_n, x_n)\}$ where all x values belong to X which is the space of input patterns. In SVR, the main goal is to find a function $f(x)$ which has that has at most ε deviation from the obtained outputs y_i for all the training data, and thus, is as flat as possible. If our errors are less than ε, this situation is can be tolerated, but otherwise not. The $f(x)$ linear function is

$$f(x)=w(x+b) \tag{10}$$

where b is the deviation vector, w is weight and x is a linear input value. Here, we can add slack variables ξ and ξ^* and we can write this problem as a convex optimization problem

$$\min=\frac{1}{2}\left\|w^2\right\|+C\sum_{i=1}^{n}(\xi_i+\xi_i^*) \tag{11}$$

subject to

$$\begin{aligned} &y_i-wx_i-b\le\varepsilon+\xi_i\\ &wx_i+b-y_i\le\varepsilon+\xi_i^*\\ &\xi_i,\ \xi_i^*\ge 0,\quad i=1,\,2,\,\ldots,\,n \end{aligned} \tag{12}$$

In equation (11), C is a constant that indicates the penalty coefficient for the purpose of set up an control mechanism on the slack variables. Furthermore, the constant $C>0$ determines the trade-off between the flatness of f and the amount up which deviations larger than ε are tolerated. ε indicates insensitive loss function.

To solve this problem, a standard dualization using Lagrangian multipliers will be into consideration;

$$\begin{aligned} L:=\frac{1}{2}\left\|w^2\right\|+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)-\sum_{i=1}^{n}(\eta_i\xi_i+\eta_i^*\xi_i^*)-\sum_{i=1}^{n}\alpha_i(\varepsilon+\xi_i-y_i+wx_i+b)\\ -\sum_{i=1}^{n}\alpha_i^*(\varepsilon+\xi_i^*+y_i-wx_i-b) \end{aligned} \tag{13}$$

in equation (13), L is a Lagrangian and $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ are Lagrangian multipliers. Dual variable in equation (13) have to satisfy positivity constraint as follow;

$$\eta_i^*,\ \alpha_i^*,\ \eta_i, \alpha_i \geq 0 \tag{14}$$

The dual Lagrangian form is

$$\max -\frac{1}{2}\sum_{i,j=1}^{n}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)\left(x_i,x_j\right)-\varepsilon\sum_{i=1}^{n}\left(\alpha_i+\alpha_i^*\right)+y_i\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right) \tag{15}$$

subject to

$$\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)=0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C] \tag{16}$$

Based on the Karush–Kuhn–Tucker theorem, regression model, and w value are expressed by:

$$w=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)x_i \tag{17}$$

$$f(x)=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)\left(x_i,x\right)+b \tag{18}$$

When data is not linear, a transformation function will be applied from one dimension to another one, so that now data is easily separable using a hyperplane. For the non-linear problem, we use the same equations (11)–(18) with some differences such as using Kernel functions. This, for example, could be achieved by simply pre-processing the training patterns by x_i mapφ. For example, in equation (11), instead of x input value, we use $\varphi\left(x_i\right)$ kernel transfer function for the inputs. Similarly, in equation (18), instead of $\left(x_i, x\right)$ statement, we use $K\left(x_i, x_j\right)$ kernel function whose value equals the inner product of two vectors, x_i and x_j, in the feature space $\varphi\left(x_i\right)$ and $\varphi\left(x_j\right)$. There are different kernel functions such as radial basis kernel, polynomial kernel, etc. In this study, the commonly used kernel functions in SVR include linear, radial basis function (RBF), polynomial, and sigmoid will be used:

- linear kernel

$$K\left(x_i,x_j\right)=x_i x_j \tag{19}$$

- radial basis kernel

$$K\left(x_i,x_j\right)=\exp\left(-a_1\left|x_i-x_j\right|\right)^2 \tag{20}$$

- polynomial kernel

$$K\left(x_i,x_j\right)=\left(a_1 x_i x_j+a_2\right)^d \tag{21}$$

- sigmoid kernel

$$K\left(x_i,x_j\right)=\tanh\left(a_1\left(x_i x_j\right)+a_2\right) \tag{22}$$

3.3.5. Random Forest Regression

Random Forest is an ensemble machine learning technique introduced by Breiman [7], which is based on decision trees. It is based on building decision trees on a training dataset to obtain its prediction results, and mean prediction of individual trees. In other words, it is an extension of the decision tree algorithm, in which each decision tree is trained and all of these trees are combined. Random Forest algorithm application steps can be defined as follows [7];

- Choose at random k data points from the training set.
- Build the decision trees associated with the selected data points as subsets.
- Choose the number N of trees you want to build.
- Repeat steps 1 and 2
- For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

3.3.6. Extreme Gradient Boosting (XGBoost) algorithm for regression

XGBoost is an acronym for the eXtreme Gradient Boosting algorithm. The XGBoost is a scalable end-to-end tree boosting system approach that uses a gradient boosting framework and is also a decision tree-based ensemble method [9, 10]. XGBoost is a powerful approach for building supervised regression models. When the XGBoost algorithm is applied to regression problems, new regression trees are continuously added and then the residuals of the previous model are fitted by the newly generated decision tree. Each decision tree calculates the feature and threshold with the best branching effect and completes the split construction.

3.4. Hyperparameters optimization

Hyperparameters are any parameters that can be set arbitrarily by the user before training begins. Machine learning models include hyperparameters that we need to set to fit the model to our data set. A model parameter is internal and cannot be interfered with, whereas hyperparameters are external and can be optimised using different approaches. Optimisation of hyperparameters plays an important role in increasing the prediction accuracy of the proposed model.

In this study, we have used random search, grid search and Bayesian optimisation algorithms for hyperparameter optimisation. Grid search is an exhaustive search through a set of manually specified hyperparameter values. Random search, on the other hand, randomly searches the grid space instead of performing an exhaustive search, i.e., it tries randomly selected combinations of parameters.

Bayesian optimization provides a technique, based on the Bayes theorem, for constructing a posterior distribution of functions that best describes the function to be optimised. Bayesian optimisation has two main features, which are (1) a Bayesian statistical model for modelling the objective function and (2) an acquisition function for deciding where to sample next. Once the objective model is determined with respect to an initial space-filling experimental design, often consisting of points chosen uniformly at random, they are used iteratively to allocate the remainder of the budget of N function evaluations, as shown in the Algorithm 1 [18].

Three optimization algorithms optimize the model parameters that are shown in Table 6 using a repeated cross-validation (rcv) technique as a performance metric. As the resampling method for the train set in this study, we have adopted cross-validation with 10-fold samples to prevent overfitting and improve the robustness of the results. Multiple linear regression does not have any hyperparameters, so we have just applied the cross-validation method. On the other hand, ridge and lasso regression methods include lambda (λ) value that is a penalty parameter as called L2-norm and L1-norm respectively. To improve the Support Vector Regression (SVR) model's accuracy, three major parameters have been taken into consideration including regularization constant (C), kernel function (radial, polynomial and linear), and gamma. C is the penalty parameter and gamma defines how far influences the calculation of a plausible line of separation. Furthermore, for SVR models based on polynomial kernel function, we have also taken into consideration scale and degree hyperparameters.

Algorithm 1 Bayesian optimization algorithm

Input: Initial points n_0, total number of iterations N
Place a Gaussian process prior on f
Observe f at n_0 points according to an initial space-filling experimental design. Set $n \leftarrow n_0$
while $n \leq N$ **do**
  Update the posterior probability distribution on f using all available data
  Let x_n be a maximizer of the acquisition function over x, where the acquisition function is computed using the current posterior distribution
  Observe $y_n \leftarrow f(x_n)$
  Increment $n \leftarrow n + 1$
end while
Return a solution: either the point evaluated with the largest $f(x)$, or the point with the largest posterior

Table 6. Hyperparameters optimization components for the regression algorithms

Algorithm	Validation	Hyperparameters	Optimum algorithm
Multiple linear regression	RCV		manual
Ridge regression	RCV	lambda	grid, random
Lasso regression	RCV	lambda	grid, random
Support Vector Regression	RCV	regularization constant (c), gamma, degree, scale and kernel function	grid, random,BO
Random Forest	RCV	mtry, ntree	grid, random, BO
XGBoost	RCV	nrounds/max-depth/gamma/colsample-bytree /min-child-weight/sub-sample/lamda/alpha	grid, random, BO

The optimization for the Random Forest is about deciding $mtyr$ and $ntree$ hyperparameters. The $mtyr$ is number of variables randomly sampled as candidates at each split, while the $ntree$ is number of trees to grow. There are nine hyperparameters in the XGBoost algorithm: max depth, min child weight, nrounds, η, γ, subsample, colsample bytree, α and λ.

3.5. Performance indicators

In this study, as a usual way of measuring the accuracy of the forecasts, mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) have been used. While it is desired that the values belonging to MAE, RMSE, and MAPE are close to zero in terms of accuracy of the models, R^2 values close to one indicate that the predicted values perfectly match actual values.

A mathematically forecasting error can be defined as follows

$$e_k = y_k - \overline{y_k} \tag{23}$$

where in equation (23), e_k is the error at the element k, y_k is the actual value and $\overline{y_k}$ is the predicted value. The four indicators are defined as follows

$$\mathrm{MAE} = \frac{1}{n}\sum_{k=1}^{n} |e_k| \tag{24}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n} e_k^2} \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} e_k^2}{y_k - y_k'} \tag{26}$$

$$\mathrm{MAPE} = \frac{100}{n}\sum_{k=1}^{n} \frac{|e_k|}{|y_k|} \tag{27}$$

The MAE, RMSE, R^2, and MAPE indicators are defined with equations (24)–(27) can enable a comparison of the regression algorithms for all datasets in terms of difference between the predicted and expected values of the hospital electricity and natural gas demands.

4. Results and discussions

In this part, six different machine learning regression algorithms have been performed on each dataset for electricity and natural gas forecasting, including four single learning models and two ensemble learning models. This section aims to present the efficiency of different supervised machine learning regression algorithms with hyperparameter optimisation over the considered contrast data-driven models for accurate hospital energy consumption forecasting. The results of each dataset will be handled separately, which were obtained using R programming software. The average value of RMSE, MAPE, R^2, and MAE accuracy metrics has been acquired after 10 experiments were considered for the performance evaluation of optimized six different regression algorithms. All the models and hyperparameters optimization were performed in a computer with AMD Ryzen 5 3500U CPU @2.10 GHz, 8 GB of RAM memory.

Although monthly electricity and natural gas consumption in healthcare facilities is considered and analyzed in general, it is clear that short-term consumption, such as daily and hourly, is very important for the predictions to be made in the context of energy management. In this context, monthly, daily and hourly electricity and natural gas consumption datasets were used to investigate the prediction performance of the proposed regression algorithms at different time periods. Input variables of each dataset are determined based on time granularity. For example, monthly consumptions are influenced by more aggregate characteristics such as number of patients, floor area, etc., whereas hourly consumption are influenced by more specific characteristics such as weather conditions, type of day or type of day period.

4.1. Results of default models

As a first step, the related analyses were carried out to find the results of non-optimized models that have default values. For example, Table 7 shows the results of the regression default models for the daily electricity dataset.

Table 7. Daily electricity dataset prediction accuracy on the training and testing datasets

Regression algorithm	RMSE [kwh]	R2	MAE [kwh]	MAPE [%]
	Training			
MLR	2,366.07	0.69	2,036.22	16.22
Ridge	2,475.65	0.67	2,113.59	17.05
Lasso	2,446.51	0.68	2,100.34	16.89
SVR-radial	2,220.74	0.73	1,804.79	14.97
SVR-poly	2,596.79	0.62	2,135.61	17.29
SVR-linear	2,393.45	0.68	2,024.45	16.42
Random Forest	1,990.56	0.79	1,731.06	14.08
XGBoost-linear	2,390.93	0.68	2,053.53	16.33
	Testing			
MLR	1,749.28	0.35	1,455.63	12.47
Ridge	1,842.11	0.32	1,548.32	13.63
Lasso	1,842.67	0.31	1,535.26	13.5
SVR-radial	1,930.58	0.33	1,634.32	14.24
SVR-poly	2,352.56	0.15	1,967.22	17.1
SVR-linear	1,885.58	0.3	1,531.97	13.31
Random Forest	1,602.99	0.48	1,333.71	11.72
XGBoost-linear	1,833.35	0.29	1,499.53	12.82
XGBoost-tree	1,833.35	0.29	1,499.53	12.82

As shown in Table 7, the Random Forest model outperforms the other eight models with an RMSE of 1602.99 kWh, MAPE of 11.72% and MAE of 1331.71 kWh for the testing set. It also outperforms by a significant margin with an RMSE of 1990.56 kWh, MAE of 1731.06 kW and MAPE of 14.08% for the training set. For both training and testing data sets, Random Forest gives the highest results in terms of R^2 with 0.79 and 0.48, respectively, where R^2 indicates that the level of predicted values matches the actual values. On the other hand, the SVR-radial method is the second most suitable method for both the testing and training sets, while SVR-poly has the worst results among all the models for both the training and test sets.

Table 8. Monthly datasets prediction accuracy on testing data

Regression algorithm	RMSE [kwh]	R2	MAE [kwh]	MAPE [%]
	Training			
MLR	61927.80	0.98	45435.17	19.34
Ridge	71285.64	0.97	59170.10	27.14
$Lasso$	45239.07	0.99	34956.08	17.42
SVR-radial	128118.80	0.89	92987.06	40.00
SVR-poly	140521.60	0.80	97713.40	36.16
SVR-linear	47867.68	0.98	37978.87	17.33
Random Forest	70475.84	0.96	55958.61	23.01
XGBoost-linear	90057.57	0.92	59414.30	19.42
XGBoost-tree	90057.57	0.92	59414.30	19.42
	Training			
MLR	76233.29	0.66	57580.65	13.63
Ridge	85768.55	0.61	65421.82	15.45
Lasso	94154.62	0.65	67770.23	14.26
SVR-radial	85913.99	0.67	70904.15	17.38
SVR-poly	70569.21	0.68	57180.74	13.85
$SVR-linear$	75420.23	0.68	56714.94	13.40
Random Forest	79873.45	0.73	63940.37	15.79
XGBoost-linear	84629.65	0.62	69929.11	17.91
XGBoost-tree	84629.65	0.62	69929.11	17.91

For monthly datasets, the results of the testing set have been evaluated as displayed in Table 8. Considering the two datasets depending on the performance indicators, the prediction results show that the prediction performance of SVR-linear on the monthly electricity dataset has the best performance values including RMSE value of 75420.23 kWh, MAE value of 56714.94 kWh, R^2 value of 0.68 and MAPE value of 13.40%.

Similarly, for the monthly natural gas testing data set, the Lasso regression has the best values with an RMSE value of 45239.07 kWh, MAE value of 34956.08 kWh, R^2 value of 0.99 and MAPE value of 17.42%. As can be seen from the results, the RMSE and MAE values are not compatible with the MAPE values, so this incompatibility is due to the size of the datasets, also the RMSE value is affected by the size of the sample and is more sensitive to outliers than the MAPE.

Finally, considering the hourly data set, the results show that XGBoost-tree is the most effective model for predicting energy consumption for the training set, giving the best values of RMSE (118.62 kWh), MAE (90.43 kWh), R^2 (0.71) and MAPE (16.91%). On the other hand, the SVR-radial model has the best results with an RMSE of 100.34 kWh, MAE of 78.69 kWh, R^2 of 0.49 and MAPE of 16.80% for the test set.

Furthermore, the computation time of the regression algorithms in the non-optimized situation was also examined. The computation times spent for model training and testing in an aggregate manner are listed in Table 9.

Table 9. Computation time of regression algorithms in default situation [s]

Regression algorithm	Monthly natural gas	Monthly electricity	Daily electricity	HourlyElectricity
MLR-default	0.083	0.051	0.152	0.099
Ridge-default	0.679	0.637	1.340	0.819
Lasso-default	0.498	0.509	1.045	0.715
SVR-radial-default	0.032	0.042	0.078	6.911
SVR-poly-default	0.021	0.017	0.050	3.419
SVR-linear-default	0.023	0.020	0.052	3.372
RF-default	0.182	0.200	0.207	9.586
XGBoost-linear-default	0.729	0.848	0.916	1.787
XGBoost-tree-default	0.729	0.797	0.987	2.300

Although computation time is a more meaningful performance indicator in hyperparameter optimisation, it is also important to state how much time is required for each model in the default situation. The computation time depends on many factors such as the size of the data, the type of model and the technical specifications of the computer. As can be seen from Table 9, the XGBoost models have much longer running times than the other supervised regression models for three datasets, while the Random Forest model has the longest running time at 9.586 seconds for the hourly dataset. The MLR model requires less computation time than other linear and non-linear techniques in the hourly dataset, while SVR-related models have minimum computation time in the remaining datasets.

4.2. Results of optimized models

Regardless of the results obtained for non-optimised hyperparameter values, the best model order may change as a result of hyperparameter optimisation. For hyper-parameter optimisation, grid search and random search were applied to Ridge and Lasso regression, while Bayesian optimisation was applied to

SVR, Random Forest and XGBoost models in addition to grid and random search. Once the hyperparameters have been optimised for the daily electricity dataset, the results obtained are shown in Table 10. Considering the tuned prediction results for the test dataset, the Ridge regression model optimised with grid search has the best values with its RMSE of 1749.28, MAE of 1455.63 kWh, R^2 of 0.346 and MAPE of 12.47% values between the individual learning algorithms. The ridge regression model optimised with grid search has reached its minimum RMSE metric at lambda value 0 as shown in Figure 3.

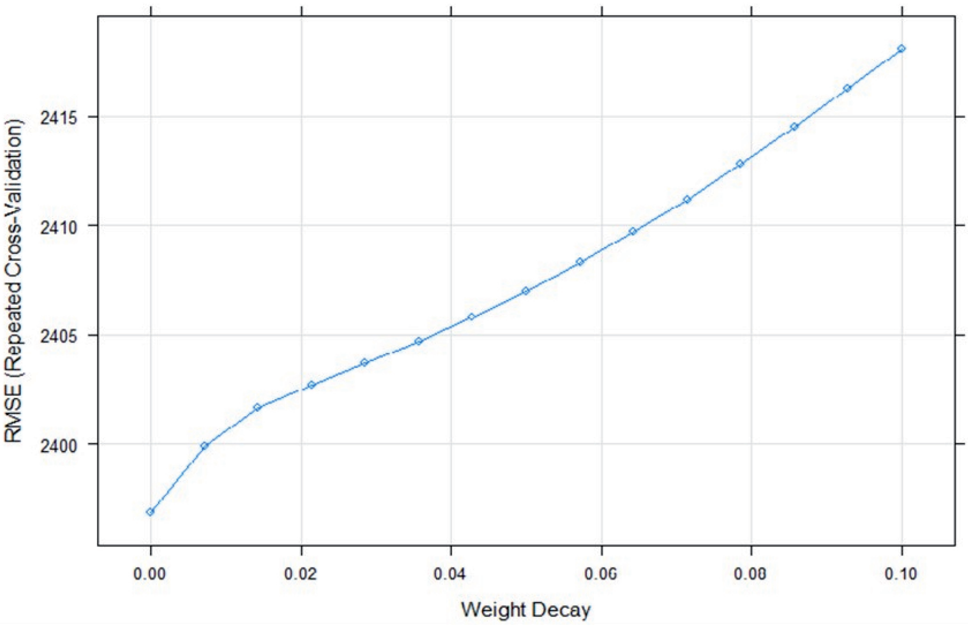

Figure 3. The RMSE and Lambda values for the optimized ridge regression

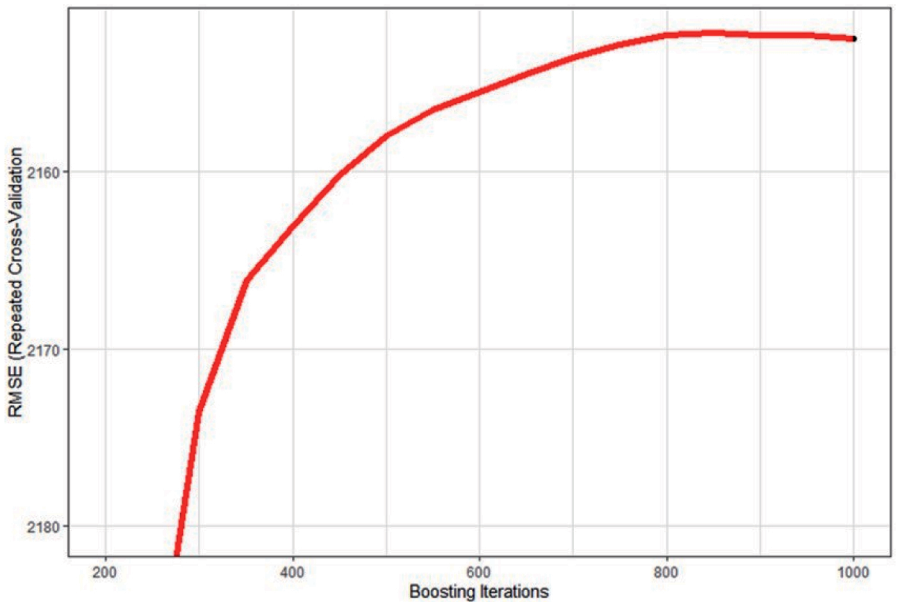

Figure 4. RMSE values and iterations for the optimized XGBoost-tree model

Among the ensemble regression algorithms, the XGBoost-tree model optimised with Bayesian optimisation has the best performance metric values with its RMSE of 1646.87 kWh, MAE of 1250.94 kWh, R^2 of 0.5 and MAPE of 10.06%. It can be concluded from the results that the XGBoost-tree model optimised with Bayesian optimisation has the best performance values among all optimised regression algorithms. In this direction, it is generally expected that ensemble learning algorithms have better performance values than single learning algorithms. As shown in Figure 4, the XGBoost-tree algorithm for the daily data set reaches its minimum RMSE value at the 800th iteration when hyperparameter optimisation is performed using Bayesian optimisation for the training set.

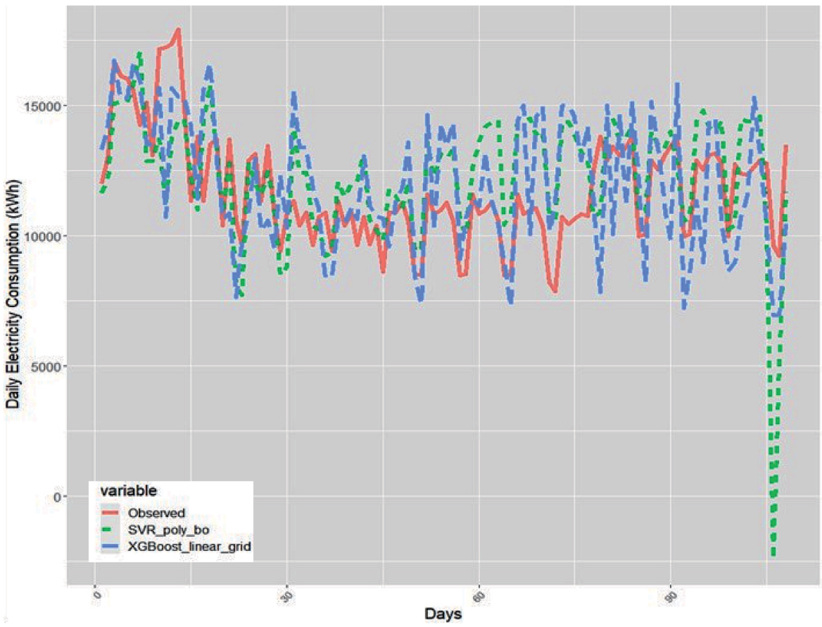

Figure 5. The least suitable two models among all optimized for testing set

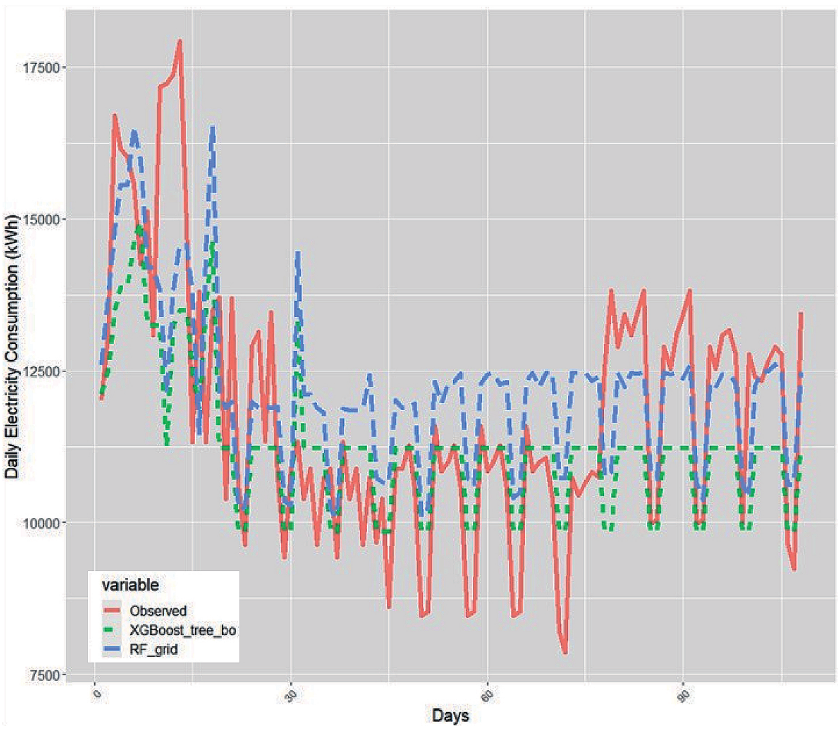

Figure 6. The best fitted two models among all optimized for testing set

Figure 5 displays two regression models SVR-poly and XGBoost-linear models optimized with Bayesian optimization and grid search, which have the worst accuracy results among all optimized models, in particular, SVR-poly model has large deviations. Figure 6 shows a comparison of the observed testing set values with the XGBoost-tree and Random Forest models, which are the best models from the point of minimum RMSE, MAE, MAPE, and maximum R^2 values. MAE indicator of XGBoost-tree model is lower compared to the MAE of Random Forest, while its RMSE is higher. In this case, the MAPE and R^2 values have been taken into consideration and these two performance indicators have better values in the XGBoost-tree model, so XGBoost-tree model optimized with Bayesian optimization has been evaluated as the best fitted model for daily electricity consumption.

Looking at the XGBoost-tree performance metrics in terms of non-optimised and optimised models, shown in Tables 7 and 10, the approximate improvements achieved were RMSE of 10.2%, MAE of 16.6%, R^2 of 72% and MAPE of 21.5%. These results show that hyperparameter optimisation is very significant in increasing model accuracy in terms of prediction. Using the same computer specifications, different regression models were compared in terms of the CPU time required to obtain the optimised model results, as shown in Figure 7 and Table 10. Among the hyperparameter optimisation methods, grid search and Bayesian optimisation require more time than random search. In particular, when grid search is applied to ensemble algorithms, quite a lot of time is spent. Looking at the computation time from the point of view of the regression models, the XGBoost-tree model optimised with grid search has a maximum computation time of 4106.97 seconds. On the other hand, the Lasso regression model optimised with random search has a minimum running time of 2.84 seconds among all optimised models. Moreover, the trade-off between computation time and accuracy can be considered as a decision point for the right model selection.

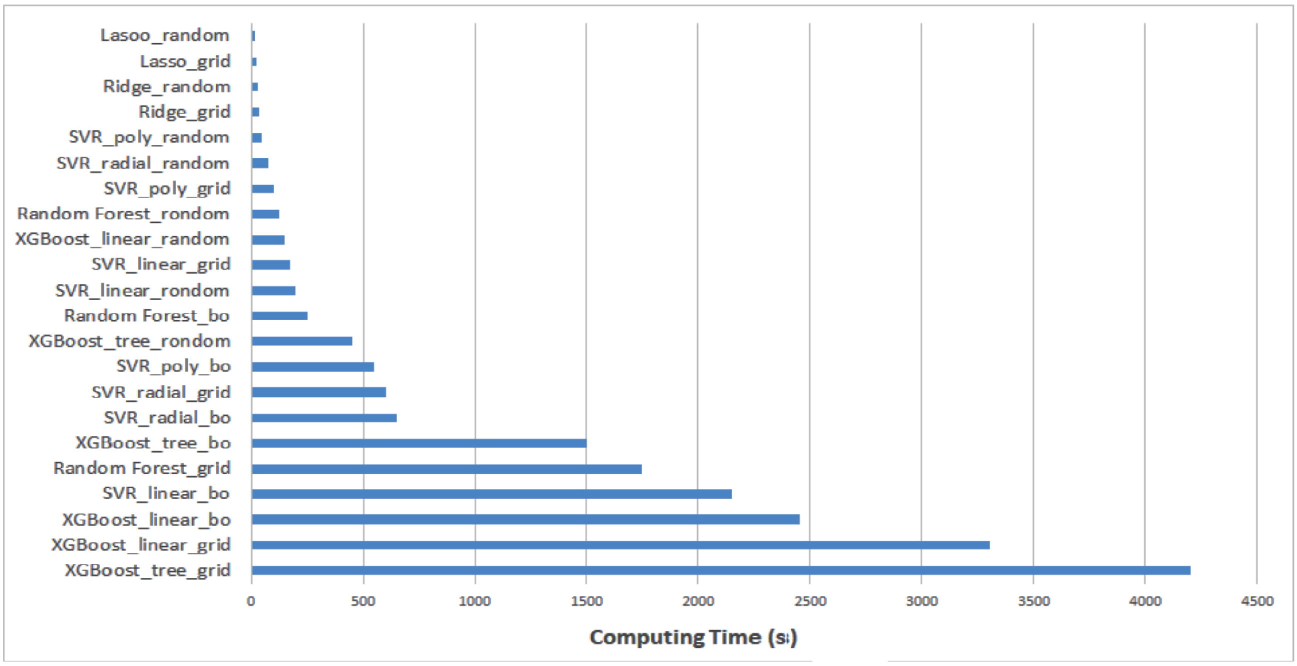

Figure 7. Computing times of the optimized algorithms for daily electricity

For example, the Random Forest model optimized by random search, which has values close to the XGBoost-tree model in terms of performance indicators, and which has a very short computation time of 45.99 seconds, can be considered as a preferable option.

All the hyperparameter optimisation steps performed for the daily electricity dataset were also applied to three other datasets, namely monthly electricity, monthly natural gas and hourly electricity. Table 11

Table 10. The results of the optimized models for daily electricity prediction

Model	Method	RMSE [kwh]	R2	MAE [kwh]	MAPE [%]	Computation time [s]
		Training				
Ridge	random search	2368.94	0.68	2038.92	16.23	
Lasso		2366.40	0.69	2036.83	16.22	
SVR-radial		2139.42	0.75	1680.46	13.7	
SVR-poly		2161.04	0.74	1717.51	13.87	
SVR-linear		2399.43	0.68	2050.35	16.59	
Random Forest		1995.00	0.79	1733.89	14.07	
XGBoost-linear		1317.67	0.9	987.11	7.48	
XGBoost-tree		1967.67	0.78	1691.49	13.39	
Ridge	grid search	2366.07	0.69	2036.22	16.22	
Lasso		2366.41	0.69	2036.84	16.22	
SVR-radial		2048.40	0.76	1625.66	12.96	
SVR-poly		2405.90	0.68	2014.32	16.49	
SVR-linear		2392.55	0.68	2045.44	16.55	
Random Forest		1994.13	0.79	1728.68	14.06	
XGBoost-linear		1204.28	0.90	676.71	4.94	
XGBoost-tree		1986.89	0.78	1697.25	13.49	
SVR-radial	Bayesian optimization	2080.47	0.76	1678.20	13.48	
SVR-poly		2132.76	0.75	1681.22	13.65	
SVR-linear		2391.99	0.68	2043.89	16.55	
Random Forest		1982.90	0.79	1720.66	13.98	
XGBoost-linear		1717.16	0.86	1400.56	10.18	
XGBoost-tree		2594.94	0.73	2189.97	15.66	
		Testing				
Ridge	random search	1766.82	0.33	1458.63	12.49	7.22
Lasso		1754.08	0.34	1456.19	12.47	2.84
SVR-radial		2177.99	0.3	1780.34	15.7	27.76
SVR-poly		2296.26	0.27	1774.18	15.61	21.41
SVR-linear		1842.21	0.32	1517.49	13.23	125.85
Random Forest		1596.08	0.48	1330.45	11.66	45.99
XGBoost-linear		2006.43	0.35	1633.48	14.25	63.45
XGBoost-tree		1658.15	0.45	1307.01	11.47	480.74
Ridge	grid search	1749.28	0.35	1455.63	12.47	11.74
Lasso		1754.14	0.34	1456.20	12.47	4.83
SVR-radial		2066.19	0.35	1674.20	14.73	554.52
SVR-poly		1883.16	0.32	1553.98	13.57	30.91
SVR-linear		1841.97	0.32	1505.41	13.11	68.39
Random Forest		1598.44	0.48	1337.22	11.73	1856.92
XGBoost-linear		2360.65	0.26	1943.72	16.87	3081.93
XGBoost-tree		1650.39	0.46	1312.21	11.43	4106.97
SVR-radial	Bayesian optimization	1972.27	0.38	1604.81	14.09	1450.27
SVR-poly		2360.26	0.26	1814.17	16.02	507.42
SVR-linear		1850.44	0.31	1506.54	13.13	2048.09
Random Forest		1595.53	0.48	1332.24	11.65	218.08
XGBoost-linear		1688.45	0.44	1367.70	11.72	2256.88
XGBoost-tree		1646.87	0.50	1250.94	10.06	1624.22

shows the best-fitting models ranked in the first two positions for three datasets. On the other hand, Figure 8 shows the observed and predicted values in terms of hourly and monthly data sets for the reference hospital based on the results listed in Table 11. The results show that when ensemble learning algorithms are considered, two of the most popular models, XGBoost and Random Forest, give the best results for constructing the energy consumption prediction model of the reference hospital.

Table 11. The results of the optimized models for the other three datasets

Energy type	Rank	Model	Method	Data	RMSE [kWh]	R2	MAE [kWh]	MAPE	Computation Time [s]
Monthly NG	1	XGBoost-tree	grid search	training	44588.05	0.99	29708.33	11.08	4185.83
				testing	44432.21	0.986	34041.24	14.88	4185.83
	2	Lasso	grid search	training	63385.4	0.97	37907.41	10.54	4.38
				testing	45081.1	0.984	31921.63	11.64	4.38
Monthly E	1	XGBoost-tree	grid search	training	34533.17	0.91	27055.88	7.4	4170.28
				testing	69528.39	0.72	55579.62	14.3	4170.28
	2	XGBoost-linear	random search	training	40.72	1	28.38	0.008	83.41
				testing	68184.01	0.75	58050.64	14.67	83.41
Hourly E	1	Random Forest	grid search	training	118.27	0.72	91.26	17.16	15212.99
				testing	101.13	0.51	79.89	17.57	15212.99
	2	XGBoost-tree	random search	training	124.74	1	0.68	96.12	985.66
				testing	68184.01	0.51	80.04	17.39	985.66

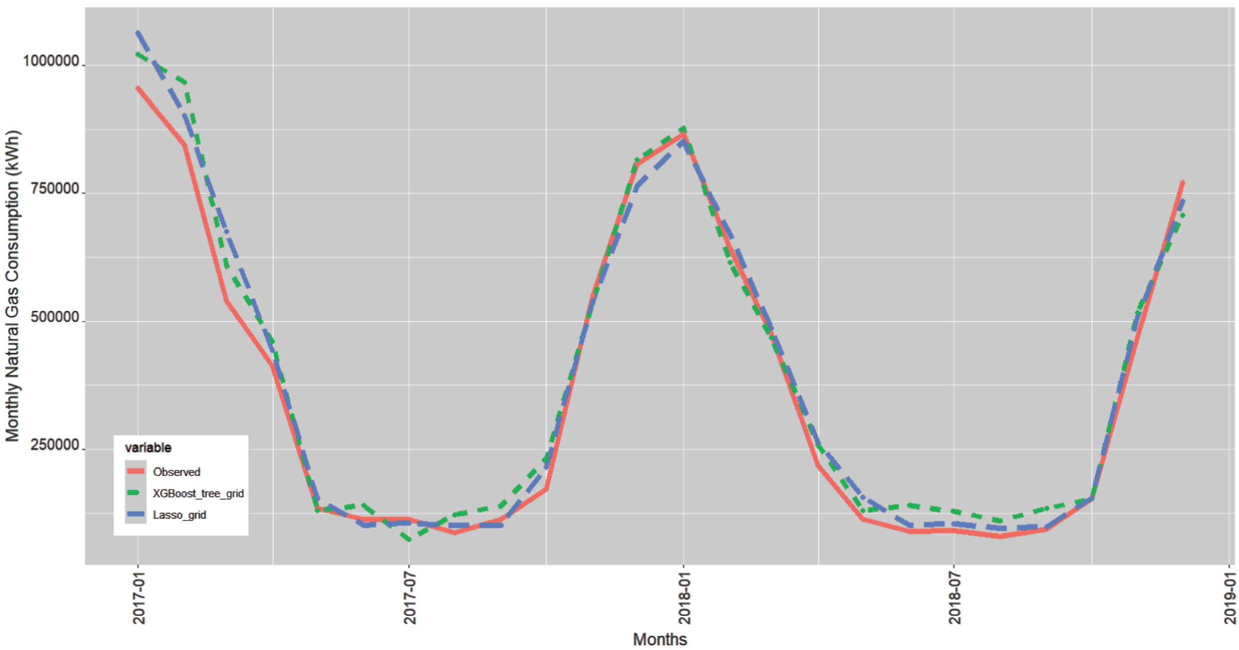

Figure 8. The best fitted models for monthly naturalgas dataset

Considering monthly natural gas consumption, the testing dataset belonging to monthly natural gas consumption reached its best accuracy metrics thanks to XGBoost-tree model optimized with grid search, where the performance indicators of this model are RMSE value of 44432.21 kWh, MAE value of 34041.24 kWh, R^2 value of 0.986, and MAPE value of 14.8%.

It reached this optimal situation with a computation time of 4185.83 seconds. On the other hand, the second best model, the lasso regression, is almost identical to the XGBoost model in terms of the test set. Considering the computation time, the lasso regression seems to be a reasonable option to predict the monthly natural gas consumption for the reference hospital.

For monthly electricity consumption data, empirical results indicate that the most accurate models are the XGBoost-tree and XGBoost-linear models, optimised with grid search and random search, respectively. Compared to the results for monthly natural gas data, the prediction models have a lower fit here, as shown in Table 11 and Figure 8. As the results of the two models are close to each other, the computation time can be evaluated as a distinguishing indicator that makes a trade-off with the accuracy for the two models mentioned. This means that the random search optimised XGBoost-linear model could be used instead of the XGBoost-tree model to predict monthly electricity.

The hourly data set with the Random Forest optimised with grid search model achieved the best accuracy indicators among all optimised models (Table 11). As mentioned earlier, the SVR-radial model outperformed all other methods in terms of all performance indicators for this dataset in the default state. Although the Random Forest model has a very good performance compared to the SVR-radial model on the training set, the performance values are quite close, especially on the test set, and even the SVR-radial model has better results on some performance metrics. Considering the running times of the Random Forest and SVR-radial models as 7606.49 and 3.84 seconds, respectively, the SVR-radial model is obviously a pretty good option for the hourly data set.

4.3. Results of the feature importance analysis

The energy consumption behaviour of healthcare buildings can be described as continuous operation and high energy use intensity, so the explanatory variables for energy consumption data sets need to be defined comprehensively. The electricity and natural gas consumption of healthcare facilities is influenced by several factors such as building structure, staff and patient intensity, weather conditions, energy systems, etc.

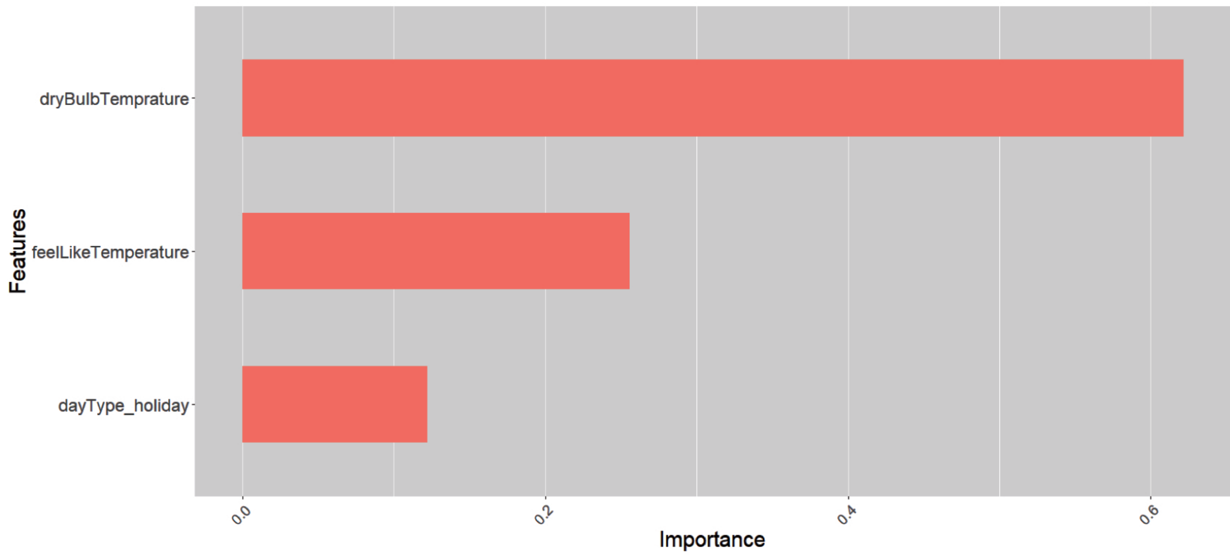

Figure 9. Feature importance analysis in XGBoost model for daily electricity prediction

However, the impact and number of these characteristics changes depending on the temporal granularity of the consumption data. Furthermore, it is obvious that these characteristics have different weights on the energy consumption. In the data preparation section, the feature selection process has been carried out and some features have been excluded from the inputs depending on the correlation analysis. This section describes the importance of the features on the target values in each dataset when used in the best-fit regression model. For example, looking at the daily data set, the daily dry bulb temperature value appears to be the most important feature, as shown in Figure 9. Healthcare facilities have HVAC systems and large chillers for cooling on hot days. As the temperature reading increases during hot months, electricity consumption increases due to the need for cooling. Therefore, dry bulb temperature has a large impact on electricity consumption, especially on hot days. Similarly, for the hourly dataset, the outdoor dry bulb temperature is the largest contributor to electricity consumption for our reference hospital, as shown in Figure 10.

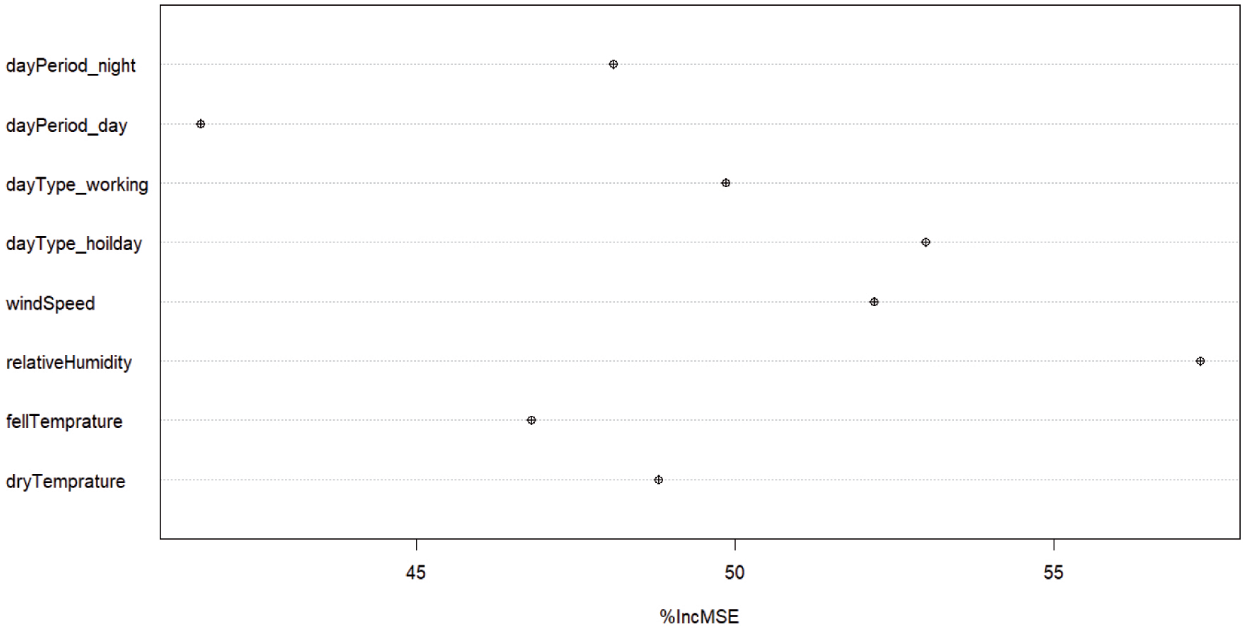

Figure 10. Feature importance analysis in Random Forest model for hourly electricity prediction

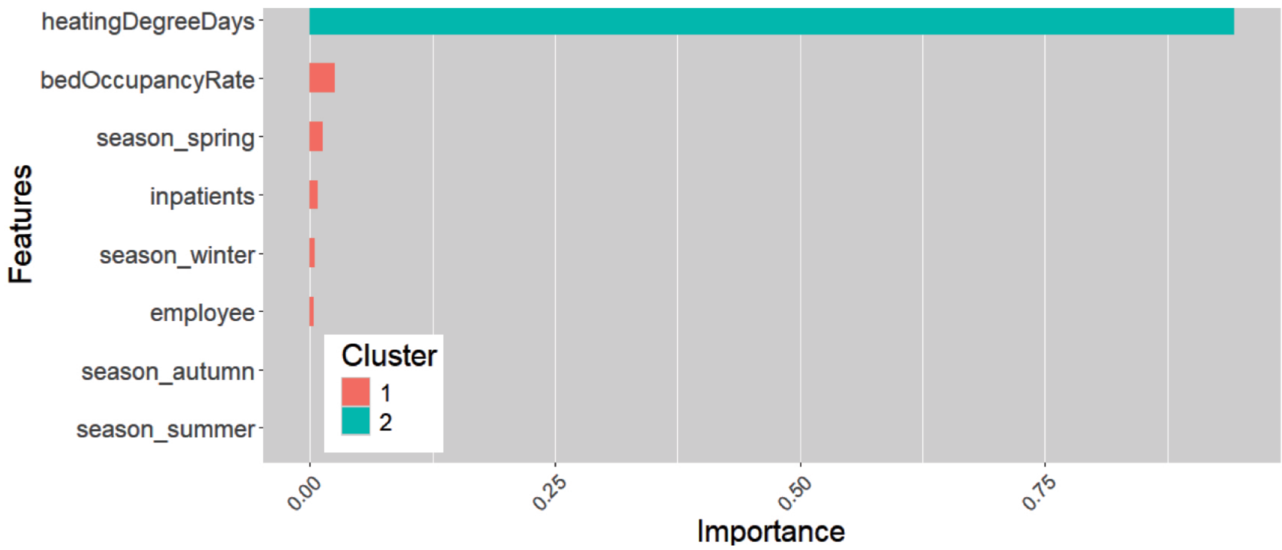

Figure 11. Feature importance analysis in XGBoost model for monthly natural gas prediction

Finally, for the monthly datasets of both natural gas and electricity, as shown in Figures 11 and 12, the most important input characteristics that have an impact on natural gas and electricity consumption are heating degree days and cooling degree days. From the importance analysis it can be concluded that weather-related input factors have a large impact on the energy consumption of healthcare facilities.

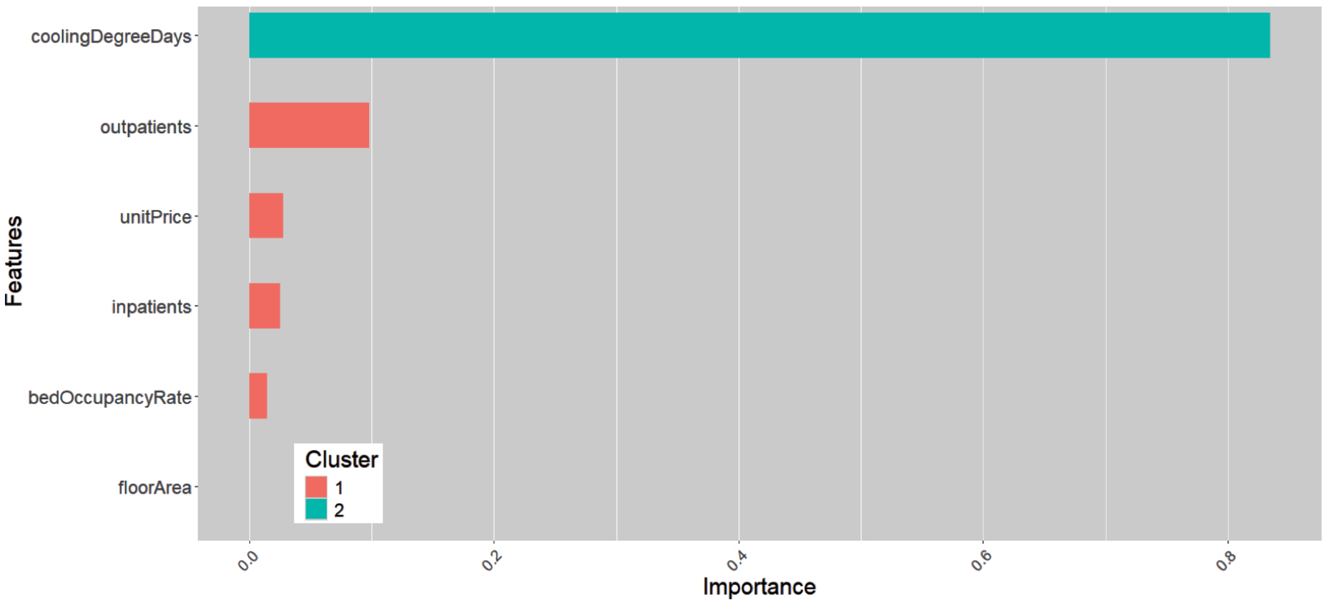

Figure 12. Feature importance analysis in XGBoost model for monthly electricity prediction

The second most important features after temperature-related features for four datasets are day type for daily electricity, relative humidity for hourly electricity, bed occupancy rate for natural gas and summer season which is a categorical variable for monthly electricity. Also, in the four datasets, some features have no effect when using the regression models specified on the target value.

5. Conclusion

In recent years, as the scope of services has expanded, the energy requirements of healthcare facilities have become one of the most important factors in terms of cost and supply methods. Energy planning in such energy-dependent buildings cannot be achieved without an understanding of past, present and future energy consumption. Therefore, accurate estimation methods of the energy load demand become crucial to determine the energy supplier selection process for the healthcare facility. By analysing the complex energy system of the healthcare facility, this study constructed various predictive models for electricity and natural gas consumption using machine learning regression algorithms (MLR, SVR, XGBoost, etc.).

This study compared six different machine learning regression models for predicting energy consumption at three different time granularities, including four single learning models and two ensemble learning models. In addition, the SVR algorithm was evaluated as three different models depending on the kernel function used, while the XGBoost algorithm was evaluated as two different models, tree and linear. In the default situation, i.e., with non-optimised hyperparameters, the empirical results indicated that the most accurate models in the single and ensemble learning categories were SVR and RF, respectively, especially for test sets. On the other hand, once the hyperparameters are optimised based on grid search, random search and Bayesian optimisation, the ensemble learning algorithms, in particular the XGBoost related models, come out on top. In addition, SVR-related models and lasso regression are notable alternative single machine learning algorithms. It was also found that the effectiveness of MLR, lasso and ridge regression decreased as the data size increased and that SVR, RF and XGBoost methods performed much better for daily and hourly data sets.

Many meteorological, calendar and occupancy variables are included to improve the quality of the predictions, and all variables are distributed across the datasets based on their time periods. Feature importance analysis was used to obtain a hierarchical ordering of the input features based on their importance to the target output.

Although the regression algorithms used perform well in forecasting electricity and natural gas demand, different approaches such as Long Short Term Memory (LSTM), reinforcement learning methods can also be used to observe the effects in a wide range. Furthermore, future research can focus on different hyper-parameter optimisation algorithms such as Genetic Algorithms, Particle Swarm Optimisation, etc. to achieve better model accuracy.

References

[1] Ahmad, T., and Chen, H. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy and Buildings 166* (2018), 460–476.

[2] Amber, K. P, Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., and Khan, M. S. Intelligent techniques for forecasting electricity consumption of buildings. *Energy 157* (2018), 886–893.

[3] Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., and Vinci, A. Electrical consumption forecasting in hospital facilities: An application case. *Energy and Buildings 103* (2015), 261–270.

[4] Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., and Zaim, S. Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. *Energy Economics 80* (2019), 937–949.

[5] Biswas, M. A. R., Robinson, M. D., and Fumo, N. Prediction of residential building energy consumption: A neural network approach. *Energy 117*, Part 1, (2015), 84-92.

[6] Braun, M. R., Altan, H., and Beck, S. B. M. Using regression analysis to predict the future energy consumption of a supermarket in the UK. *Applied Energy 130* (2014), 305–313.

[7] Breiman, L. Random Forests. *Machine Learning 45*, 1 (2001), 5–32.

[8] Cao, L., Li, Y., Zhang, J., Jiang, Y., Han, Y., and Wei, J. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Reports 6* (2020), 2751–2767.

[9] Chen, T., and Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA, 2016), B. Krishnapuram, M. Shah, A. J. Smola, C. Aggarwal, D. Shen, and R. Rastogi, Eds., ACM, pp. 785–794.

[10] Chen, T., He, T., Benesty, M., and Khotilovich, V. (2019). xgboost: Extreme Gradient Boosting (accessed on 25 July 2024).

[11] Ciulla, G., and D'Amico, A. Building energy performance forecasting: A multiple linear regression approach. *Applied Energy 253* (2019), 113500.

[12] Costa, O. L. V., de Oliveira Ribeiro, C., Ho, L. L., Rego, E. E., Parente, V., and Toro, J. A robust least square approach for forecasting models: an application to Brazil's natural gas demand. *Energy Systems 11* (2020), 1111–1135.

[13] Dulce, E., and Martinez-de Pison, F. J. *Parsimonious modeling for estimating hospital cooling demand to reduce maintenance costs and power consumption* In *Hybrid Artificial Intelligent Systems, 14th International Conference, HAIS 2019, León, Spain, September 4–6, 2019, Proceedings* (Cham, 2019), H. Pérez García, L. Sánchez González, M. Castejón Limas, H. Quintián Pardo and E. Corchado Rodríguez, Eds., vol. 11734 of *Lecture Notes in Computer Science* series, Springer, pp. 181–192.

[14] D'Amico, A., Ciulla, G., Tupenaite, L., and Kaklauskas, A. Multiple criteria assessment of methods for forecasting building thermal energy demand. *Energy and Buildings 224* (2020), 110220.

[15] Eckelman, M. J., and Sherman, J. Environmental impacts of the U.S. health care system and effects on public health. *PLOS ONE 11*, 6 (2016), e0157014.

[16] EuropeanCommission (17 February 2020) In focus: Energy efficiency in buildings (accessed on 5 march 2022).

[17] Fathi, S., Srinivasan, R., Fenner, A., and Fathi, S. Machine learning applications in urban building energy performance forecasting: A systematic review. *Renewable and Sustainable Energy Reviews 133* (2020), 110287.

[18] Frazier, P. I. *A Tutorial on Bayesian Optimization*, 2018. Working paper version avaiable from arXiv: https://arxiv.org/abs/1807.02811.

[19] González-Domínguez, J., Sánchez-Barroso, G., García-Sanz-Calcedo, J., and de Sousa Neves, N. Cox proportional hazards model used for predictive analysis of the energy consumption of healthcare buildings. *Energy and Buildings 257* (2022), 111784.

[20] Gordillo-Orquera, R., Lopez-Ramos, L. M., Muñoz-Romero, S., Iglesias-Casarrubios, P., Arcos-Avilés, D., Marques, A. G., and Rojo-Álvarez, J. L. Analyzing and forecasting electrical load consumption in healthcare buildings. *Energies 11*, 3 (2018), 493.

[21] Houimli, R., Zmami, M., and Ben-Salha, O. Short-term electric load forecasting in Tunisia using artificial neural networks. *Energy Systems 11*, 2 (2020), 357–375.

[22] Hwang, J., Suh, D., and Otto, M.-O. Forecasting electricity consumption in commercial buildings using a machine learning approach . *Energies 13* (2020), 5885.

[23] Iglewicz, B., and Hoaglin D. C. *How to Detect and Handle Outliers*. ASQC Quality Press, 1993.

[24] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. Springer, New York, 2013.

[25] Karabiber, O. A., and Xydis, G. Forecasting day-ahead natural gas demand in Denmark. *Journal of Natural Gas Science and Engineering 76* (2020), 103193.

[26] Katsatos, A. L., and Moustris, K. P. Application of artificial neuron networks as energy consumption forecasting tool in the building of Regulatory Authority of Energy, Athens, Greece. *Energy Procedia 157* (2019), 851–867.

[27] Kavousi-Fard, A., Samet, H., and Marzbani, F. A new hybrid modified firefly algorithm and Support Vector Regression model for accurate short term load forecasting. *Expert Systems with Applications 41*, 13 (2014), 6047–6056.

[28] Li, K., and Zhang, T. A novel grey forecasting model and its application in forecasting the energy consumption in Shanghai. *Energy Systems 12*, 2 (2021), 357–372.

[29] Liu, T., Tan, Z., Xu, C., Chen, H., and Li, Z. Study on deep reinforcement learning techniques for building energy consumption forecasting. *Energy and Buildings 208* (2020), 109675.

[30] Luo, X. J., and Oyedele, L. O. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. *Advanced Engineering Informatics 50* (2021), 101357.

[31] Moriñigo-Sotelo, D., Duque-Pérez, O., García-Escudero, L. A., Fernández-Temprano, M., Fraile-Llorente, P., Riesco-Sanz, M. V., and Zorita-Lamadrid, A. L. Short-term hourly load forecasting of a hospital using an artificial neural network. *Renewable Energy and Power Quality Journal 9*, 1 (2011), 441–446.

[32] Mui, K. W., Satheesan, M. K., and Wong, L. T. Building cooling energy consumption prediction with a hybrid simulation approach: Generalization beyond the training range. *Energy and Buildings 276* (2022), 112502.

[33] Neto, A. H., and Fiorelli, F. A. S. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings 40*, 12 (2008), 2169–2176.

[34] Riedmiller, M., and Braun, H. RPROP - a fast adaptive learning algorithm. In *Proceedings of the International Symposium on Computer and Information Science VII* (Antalya, Turkey, 1992), pp. 279–286.

[35] Smola, A. J., and Schölkopf, B. A tutorial on Support Vector Regression. *Statistics and Computing 14*, 3 (2004), 199–222.

[36] Somu, N., M R, G. R., and Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy 261* (2020), 114131.

[37] Somu, N., M R, G. R., and Ramamritham, K. A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews 137* (2021), 110591.

[38] Song, Y., Xie, H., Zhu, Z., and Ji, R. Predicting energy consumption of chiller plant using WOA-BiLSTM hybrid prediction model: A case study for a hospital building. *Energy and Buildings 300* (2023), 113642.

[39] Tan, Z., De, G., Li, M., Lin, H., Yang, S., Huang, L., and Tan, Q. Combined electricity-heat-cooling-gas load forecasting model for integrated energy system based on multi-task learning and least square support vector machine. *Journal of Cleaner Production 248* (2020), 119252.

[40] Timur, O., Zor, K., Çelik, Õ., Teke, A., and İbrikci, T. Application of statistical and artificial intelligence techniques for medium-term electrical energy forecasting: A case study for a regional hospital. *Journal of Sustainable Development of Energy, Water and Environment Systems 8*, 3 (2020), 520–536.

[41] Tran, D.-H., Luong, D.-L., and Chou, J.-S. Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings. *Energy 191* (2020), 116552.

[42] Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[43] Vapnik, V. N. *Statistical Learning Theory*. Wiley, 1998.

[44] Wang, L., Lee, E. W. M., and Yuen, R. K. K. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. *Applied Energy 228* (2018), 1740–1753.

[45] Yuan, J., Farnham, C., Azuma, C., and Emura, K. Predictive artificial neural network models to forecast the seasonal hourly electricity consumption for a University Campus. *Sustainable Cities and Society 42* (2018), 82–92.

[46] Zhao, J., and Liu, X. A hybrid method of dynamic cooling and heating load forecasting for office buildings based on artificial intelligence and regression analysis. *Energy and Buildings 174* (2018), 293–308.