BUILD SIMUL (2019) 12: 1033–1045
https://doi.org/10.1007/s12273-019-0552-2

A hybrid energy management approach for home appliances using climatic forecasting

Jasmeet Kaur (✉), **Anju Bala**

Computer Science and Engineering Department, Thapar University, India

Research Article

Building Thermal, Lighting, and Acoustics Modeling

Abstract

Energy management refers to saving the power by employing effective monitoring and control strategies. The demand for energy is rising in all sectors, such as residential, industrial, transportation, and agriculture, owing to our dependency on electronic appliances. Hence, of late, energy management in the households has become a pertinent issue. Electricity consumption depends on various factors, including climatic conditions, number of occupants in the household and their behavior, usage of appliances, etc. The utilization of electronic appliances and the climate conditions are inter-related; if the outside temperature is high, the usage of ACs in the house increases and vice-versa. Therefore, the climatic conditions are the most relevant factors in energy consumption. There is a need to manage the energy demand by using certain optimization approaches and by predicting the demand based on different climatic conditions. In this study, the prediction model of Artificial Neural Network (ANN) was merged with the optimization methods such as Particle Swarm Optimization (ANN-PSO) and Artificial Bee Colony (ANN-ABC). The experimental results revealed that ANN-ABC performed in a superior manner by reducing the energy consumption by up to 41.12 kWh per day. Finally, the prediction results were compared with the existing model and it was verified that the energy prediction with climatic conditions gave the better results of the power usage in the households.

Keywords

energy forecasting,
particle swarm optimization,
artificial bee colony,
artificial neural network

Article History

Received: 8 May 2018
Revised: 21 April 2019
Accepted: 24 April 2019

1 Introduction

The escalating level of energy consumption for households appliances is becoming a serious issue these days. The use of energy is increasing in the industrial, residential, transport, and commercial sectors with the passing years. Energy consumption is governed by various factors such as climate, behavior of individuals, type of appliances used in the household, routines followed by individuals to carry out their daily activities, etc. With the advent of technological advancements, there has been an increase in the residential energy consumption as the use of electric appliances has tremendously increased. The enhanced usage of such appliances has also resulted in the emission of greenhouse gases and air pollutants. Although, the usage of fluorescent lights and the management of appliances during the peak period have been initiated by the residents (Godina et al. 2016), the rising use of other kinds of appliances such as dishwashers, refrigerators, washing machines, etc. has led to an increase in the electricity demand. To reduce the residential electricity demand, energy suppliers need to forecast the demand in dwellings and supply the accurate amount. The user behavior plays an imperative role in electricity consumption and energy demand.

The usage of electricity in homes and the consumption of energy depend on the climatic conditions, and the patterns in which the appliances are used to execute the daily routines. For example, the occupant may like to watch TV while cooking and simultaneously use the dishwasher for washing the utensils. Hence, owing to the inter-dependencies of the appliances, their time of use is inter-related. The relationship between the appliances and their time of utilization is important for fulfilling the energy needs of the households (Singh and Yassine 2017). It is becoming a grueling task for the energy supplier to provide sufficient power to the residential sector at all times; hence, there is a need to forecast and optimize the energy demand. The reason behind optimizing

E-mail: jkaur3_me16@thapar.edu

the energy demand is that the suppliers can furnish the minimum energy that is required by the residents to carry out their daily activities comfortably. Energy optimization also facilitates a reduction in the release of greenhouse gases and air pollutants. Further, optimization plays a crucial role in minimizing the utility bills and the adverse climatic effects of increased appliance usage.

The two common approaches for predicting the energy demands in the residential sector are the top-down and bottom-up strategies (Swan and Ugursal 2009). Top-down approach involves predicting the demand on the basis of total power consumption, while the bottom up approach deals with the prediction at the appliance level. In this paper, the usage patterns of appliances by the occupants were investigated by means of the principal component analysis (PCA), and different clusters were formed according to it. These clusters were integrated with climatic conditions. Then, various machine learning algorithms were considered, and the most appropriate one was applied for the optimization. The energy optimization was performed by using the hybrid approach with the optimal machine learning algorithm. Such prediction and optimization can be useful for the energy supplier to gauge the demand for electricity in the residential sector.

1.1 Motivation

– The usage of electronic appliances has increased. Thus, it has become difficult for the energy suppliers to meet the demands of the dwellings.
– Greenhouse gas emissions, which have great impact on the climate, have increased with the usage of appliances.
– The optimization of energy demand is direly necessitated for the energy suppliers to meet the requirements of the dwellings so that they can perform their daily basic activities without any hindrance.

The paper is organised as follows: Section 2 overviews about the existed literature related to prediction and optimization of energy. Further, Section 3 describes the proposed methodology in detail. In Section 4, experimental results are discussed and Section 5 validates the proposed model with the existing model. Lastly, Section 6 concludes the paper.

2 Related work

There have been several researches on energy management and consumption, which have focused on different types of energy such as solar energy, wind energy, etc. Studies on load forecasting/electricity forecasting have been carried out in different sectors such as residential, buildings, offices, transportation, agriculture, etc. Many machine learning algorithms have been applied on classification and regression data to achieve the desired accuracies; optimization methods have also been discussed and implemented in many research endeavors. In this section literature pertaining to energy prediction in various sectors and different optimization techniques has been reviewed.

Wang et al. (2015) proposed hybrid renewable energy system (HRES), which comprised of solar, wind and diesel generator as a backup resource as well as battery storage for optimal operation. But the researchers have not utilized machine learning approaches for predicting power. Candanedo et al. (2017) forecasted the energy on the basis of different appliances using predictive models such as multiple linear regression (MLR), support vector machine with radial kernel (SVM), gradient boosting machine (GBM), and random forest (RF). All the models were compared with different evaluation parameters such as RMSE, accuracy, and R^2. The results indicated that gradient boosting performed better than the rest of the models when weather conditions were taken into consideration. Wang et al. (2018b) used the random forest (RF) model for predicting the energy requirements in buildings on an hourly basis, and compared it with different predictive models such as regression tree (RT) and SVR. The author also predicted the energy needs of the educational sector by determining the important factor and concluded that semester-based energy usage will be beneficial for making predictions in the residential sector rather than the annual basis. Wei et al. (2018) analysed the prevailing energy prediction models and the classification techniques for energy predictions. Numerous predictive models such as ANN, SVM, statistical regression, decision tree, GA and clustering models such as K-means clustering were discussed. Kaur and Bala (2018) predicted the energy of household by using different machine learning algorithms such as ANN, random forest, SVR, K-nearest neighbour regression (Knnr), etc. and compared the models with parameters accuracy to evaluate the better model for energy prediction of household. Divina et al. (2018) predicted the short term electricity consumption with the help of the ensemble approach. The authors used the base models of ANN, random forest (RF), and GBM, and then predicted the energy based on the ensemble approach using the three algorithms. The energy demand in smart grids by using the conventional neural network (CNN) approach and optimized the energy usage with neural network based particle swarm optimization (NNPSO) and genetic algorithm (NNGA) have been predicted by Muralitharan et al. (2018). The investigators concluded that for short-term forecasting the NNGA was more accurate but for long-term, NNPSO was ideal. Dong et al. (2016) examined the hybrid model approach for predicting energy using five data-driven methods (ANN, SVM, LSSVM, GPM, and GMM) and physics based models. However, the authors have not utilized optimization techniques to optimize the power of home appliances. Yin and Chao (2018) proposed the energy demand method, that is, multi-predictor method. The cyber swarm optimization was used for finding the optimal parameters for the predictor individually. The results were compared by using the parameters such as MSE and MAPE. Do and Cetin (2018) examined the different energy and non-energy data and discussed about the various predictive models for energy demand, including ANN, SVM, genetic programming, and Bayesian network.

The energy needs of the buildings have been predicted (Zhu et al. 2018) by using weather forecasting. Different predictive models, such as ANN, SVM, and genetic algorithm were employed by the authors for energy prediction. Wang et al. (2018a) proposed the energy prediction of the institutional building by using the ensemble approach, that is, the ensemble bagging tree (EBT) model. The EBT was compared with CART, the single prediction model for comparing the accuracy and stability of both the methods. It was concluded that EBT was superior to the single prediction model for predicting the energy demand of the building. The short-term electrical demand has been envisaged (Torabi et al. 2019) by using the hybrid approach (CBA-ANN-SVM) along with ANN and SVM. The hybrid approach of CBA-ANN-SVM (1.297 (3 clusters)) reduced the error rate as compared with the ANN (1.790) and SVM (2.015). Khakimova et al. (2017) presented a design of model predictive control system and suggested to reduce the complexity and execution time which can be accomplished by using optimization techniques. Kaur and Bala (2019) analyzed the different machine learning techniques based on the energy prediction of households and compared all the methods based on cost, power, etc. for optimizing the energy of home appliances. Rasheed et al. (2016) applied the binary multiple knapsack optimization technique to reduce the electricity bills without affecting the user comfort. The optimization was carried out on three types of appliances in response to behaviour, weather conditions and electricity prices. The results were verified by simulation of the optimization findings. The behaviour of the occupants towards the usage of different household appliances was investigated by Singh and Yassine (2017) for identifying the energy consumption patterns. An unsupervised approach was implemented for incremental data mining by applying frequent pattern mining to the energy consumption data. Lazos et al. (2014) reviewed the numerous methods of energy forecasting and minimization in the commercial buildings and the effect of weather conditions on the energy usage for the buildings.

Cottone et al. (2015) proposed the method for optimizing the energy of appliances by recognizing the user activities, which were extracted by information theory approach. Further, the knapsack optimization problem was defined for optimizing the energy needs of the households. Tian et al. (2014) applied the ordinary least squares (OLS) and spatial regression analysis in the domestic sector for discerning the requirements of the urban areas. Regression analysis and Lagrange multiplier statistical test were performed on electricity and gas usage based on the council tax and domestic energy consumption. The simulation was done to indicate the regression analysis performance. The realistic scheduling mechanism has been purposed by Mahmood et al. (2016) for classifying the appliances according to their time of use (TOU) and on the basis of different constraints. The binary particle swarm optimization (BPSO) was implemented for deriving the appliance utility and cost effectiveness. Lin et al. (2008) used particle swarm optimzation for feature selection and parameter determination of the support vector machine.

Li et al. (2017) developed an optimization method and compared it with different existing multi-objective algorithms. The techniques of GenOpt and artificial neural network (ANN) were performed, and the results were compared with non-dominated sorting genetic algorithm (NSGA-II), multi-objective particle swarm optimization (MOPSO), multi-objective genetic algorithm (MOGA) and multi-objective differential evolution (MODE), among which MODE yielded the optimum result. A hybrid decision support system have been proposed by Juan et al. (2010) using genetic algorithm (GA) and zero one goal programming model (ZOGP) for improvement of energy performance in sustainable offices. Subbiah et al. (2017) built the energy demand model for appliance usage as per the activity done by the occupant and calculated the energy consumption based on the various constraints such as appliances rating, duration of the appliance use, and the type of activity done. The method was based on individual modelling approach and different datasets were used to achieve the goals. Various machine learning algorithms were used for the prediction, and the results were compared in terms of several evaluation parameters (Gupta et al. 2017). The prediction of solar power was performed by Sharma et al. (2011) by considering weather forecast with the aid of the machine learning algorithm. Different regression algorithms were employed and the obtained results were compared based on the basis of their accuracy. The SVM-based prediction achieved the highest accuracy among all the models, that is, it was 27% more accurate than the other models. Ha et al. (2012) investigated the global energy management problem of dwellings using dynamic predictive control system. The heuristics and optimization approaches can be used for further enhancement. Ardakani and Ardehali (2014) forecasted the electric energy consumption using optimized and artificial neural network (ANN). The forecasting model employed was multi-variable regression (MVR) and ANN; the results were optimized by using particle swarm optimization (PSO) and improved PSO (IPSO). The final results indicated that IPSO-ANN performed better than PSO, with MAPE of 1.94 and 1.51 respectively for both the datasets that were utilized for forecasting. Ahmad et al. (2014) compared the ANN and SVM models with hybrid models for building energy forecasting and the results were evaluated using RMSE and correlation coefficient. Table 1 overviews different models used by some researchers for energy forecasting and minimization in the dwellings and the commercial buildings.

Based on the literature review, the followings are the key issues which need to be addressed.

- Most of the existing work is based on the accumulated energy rather than focusing on individual home appliances using clustering. So, there is a dire need to evaluate the usage patterns of individual home appliances.
- Some of the authors have predicted energy using machine learning models such as SVR, ANN, RF, Knnr, DT, etc., but none has focussed on the individual home appliance with climatic conditions.
- The hybrid optimization technique applied by the authors has not taken the climatic conditions into consideration for energy optimization along with machine learning model for the households by analysing the pattern of appliances according to climate.

Thus, the subsequent section illustrates the contributions to handle the above key issues.

2.1 Research contribution

- To analyze the energy consumption patterns for individual home appliances, principal component analysis (PCA) has been done using *K*-means clustering, whereas the existing works do not evaluate individual home appliances using clustering and PCA based on climatic conditions (Candanedo et al. 2017; Wang et al. 2018b).
- To predict the energy consumption using climatic conditions, various machine learning models have been utilized to select the best model in terms of maximum accuracy and minimum error in comparison to existing works discussed by a few authors (Divina et al. 2018; Dong et al. 2016).
- To optimize the energy consumption of home appliances, the best predicted model has been integrated with optimization techniques as to reduce the average energy consumption in contrast to others. The existing works do not consider climatic conditions through hybrid optimization techniques based on clustering (Muralitharan et al. 2018; Zhu et al. 2018; Ardakani and Ardehali 2014).

Table 1 Different models for energy forecasting and optimization

Author	Models
Candanedo et al. 2017	MLR, SVM, GBM, RF
Wang et al. 2018b	RF, RT, SVR
Divina et al. 2018	ANN, RF, GBM
Muralitharan et al. 2018	CNN, NNPSO, NNGA
Dong et al. 2016	ANN, SVM, LSSVM, GPM, GMM
Do and Cetin 2018	ANN, SVM, genetic programming, Bayesian network
Zhu et al. 2018	ANN, SVM, Genetic algorithm
Ardakani and Ardehali 2014	MVR, ANN, PSO, IPSO
Li et al. 2017	ANN, GenOpt, NSGA-II, MOPSO, MOGA, MODE
Mahmood et al. 2016	BPSO
Tian et al. 2014	OLS, regression analysis, Lagrange multiplier statistical test
Juan et al. 2010	GA, ZOGP
Torabi et al. 2019	ANN, SVM, CBA-ANN-SVM

3 Methodology

The procedure for the proposed work has been depicted in Fig. 1. The model was divided into two modules. First, the smart home data with climatic variables were integrated and different regression machine learning algorithms were tested. The second module describes the optimization technique by selecting the suitable machine learning algorithm to satisfy the energy demands of the end users without any hurdles.

3.1 Data description

The smart home data were retrieved from the almanac of

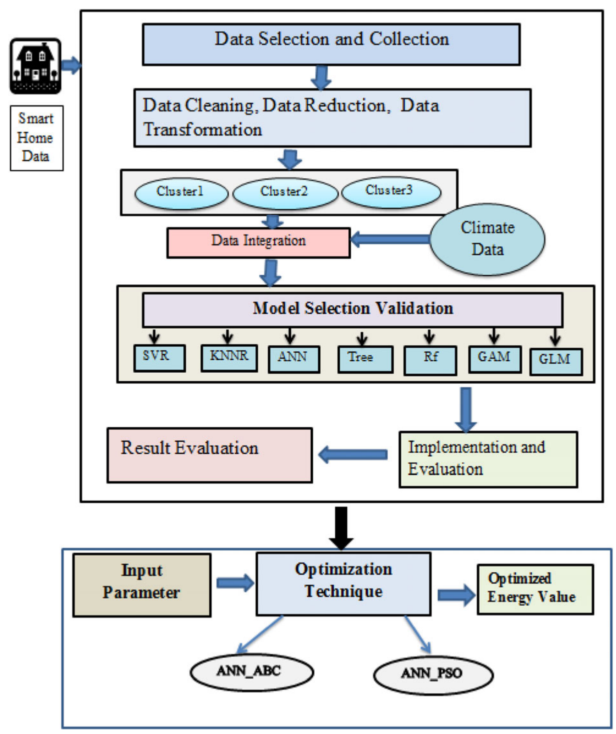


Fig. 1 Methodology for energy prediction and optimization

minutely power data set (AMPds) (Makonin et al. 2016). The data set consisted of records from one home in Vancouver, Canada from April 1, 2012 to March 31, 2014. The smart home data from different appliances were cleaned and processed. The time interval of one min was processed to an interval of one hour. Hence, the principal component analysis (PCA) was performed on smart home data pertaining to different appliances at a time interval of one hour. The main motive behind performing the PCA was dimensionality reduction based on the pattern in which the appliances were used in the household at a particular time interval, and different cluster formations of the appliances were obtained by using the *K*-means. The various cluster formations according to the usage pattern were integrated with the climate data set which involved the same time period as that of the smart home data set. Different machine learning regression algorithms were applied on all the three cluster formations, which were integrated with the climatic variables such as the temperature, wind speed, humidity, visibility, etc. All the algorithms were run on the R Tool, and different interpretations were derived on the basis of residuals graphs, accuracy, mean square error, and coefficient of determination. Different bar plots and graphs were drawn to ascertain the relationship between appliances usage and climate variables.

3.2 Models description

Different regression models, such as support vector regression (SVR), *K*-nearest neighbour regression (Knnr), random forest, tree, artificial neural network (ANN), general additive model using splines (GAMs), and multivariate adaptive regression splines (MARS) were implemented during training and testing.

3.2.1 Support vector regression (SVR)

The SVR was employed for maintaining the features that were used as predictors for model training and testing. The electricity data were nonlinear data; hence, the SVR was used for prediction as the basic principle of SVR is to map the nonlinear data and decrease the prediction risk (Chen et al. 2017). The main purpose of the method is to map the data from the training dataset to the feature space, an optimized hyperplane, by formulating the nonlinear relationship between the response and predictor variables. The basic SVR function is formulated as:

$$f(y) = \gamma \times \Theta(y) + \beta \tag{1}$$

where $\Theta(y)$ is the feature that is mapped to the input data y, and γ and β are the coefficients.

3.2.2 K-nearest neighbour regression

It was started in 1970s for pattern recognition as a non-parametric technique. It predicts the target based on some distance of the *K*-nearest neighbours. The parameters are calculated by the knn.reg() function for the prediction with different parameters tuning for achieving the good accuracy.

3.2.3 Random forest

The random forest has categorized as ensemble-learning models. It combines the different regression trees and its root node represents the different path to variables with highest importance. It selects the different samples and draws the relationship between the response and predictor variables. The basic motive of random forest is to deal with the large number of values. As the electricity data has very large values so random forest has been utilized as the prediction model for predicting energy demands.

3.2.4 Tree

The decision tree was used by the rules for the series completion of the data. They were used to decide the variables path to achieve the required results for the prediction. The prediction is based on the probability the predictor variable with the response variable. The energy demand deals with the different patterns in the households by the occupants, therefore, the decision tree has been widely used prediction model.

3.2.5 Artificial neural network (ANN)

ANN has been widely used to extricate the information from the climatic data as climatic data consist of many hidden patterns. It consists of three different layers i.e. input layer, hidden layer. By considering the previous researches, it was concluded that ANN has been widely used for predicting energy consumption as it has the ability to accommodate nonlinear data with different consumptions patterns and achieve high accuracy (Biswas et al. 2016). The Nnet package was used for training the model for predicting the energy consumption with different climate variables as an input.

3.2.6 General additive model using splines (GAMs)

GAM is the generalized version of linear models in which the predictors rely linearly or nonlinearly on smooth nonlinear functions like splines, polynomial, and step functions, etc. In this paper, we have used splines to fit smooth linear function on the wreath of predictors (x_1, x_2,..., x_p). In Eq. (2) $y_i/F(x)$ is the regression function on different predictors. ϵ_i is the noise in the different nonlinear functions and $f_j(x_{ij})$ are the different nonlinear functions on the predictors (x_p) where p is the number of different variables on which the regression function was calculated (Wood 2001) where $\sum_{j=1}^{p} f_j(x_{ij}) = f_1(x_{i1}) + f_2(x_{i2}) + \cdots f_p(x_{ip})$, i.e., the sum of different functions used on different predictors. We have used gam() function in R through splines with an approach back-fitting.

$$y = F(x) = \sum_{j=1}^{p} f_j(x_{ij}) + \epsilon_i = f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) \tag{2}$$

3.2.7 *Multivariate adaptive regression splines (MARS)*

It was introduced by Jerome H. Friedman in 1991 (Balshi et al. 2009). MARS is a generalization of stepwise linear regression which takes the form of an expansion in splines basis functions. The basis functions as well as the variables associated with each functions are automatically determined by the data. In Eq. (3) $\beta_i(x)$ is the weighted sum of the basis functions in which one basis function is multiplied by its coefficient.

$$F(x) = \sum_{j=0}^{k} \Theta_i \beta_i(x) \tag{3}$$

where $\beta_i(x)$ is the basis function which can be of three types and Θ_i is a constant coefficient.

3.3 A hybrid energy management techniques

The optimization approach has been used with the earmarked machine learning algorithm. As ANN resulted in maximum accuracy and minimum error rate as shown in Section 4, it was utilized along with the optimization techniques to reduce the energy consumption. The hybrid optimization approach of ANN-PSO, and ANN-ABC was implemented.

3.3.1 *Artificial neural network*

For electricity forecasting, ANN has been widely used because of its capability to deal with nonlinear data. This paper also involved data based on climatic conditions and appliances usage. For climatic indicators as input data, electricity forecasting was defined as a function of different climatic variables, such as temperature, humidity, visibility, wind speed etc., as inferred from Eq. (4).

$$\begin{aligned} E(t) = f(&C_{\text{temperature}}(t), C_{\text{Humidity}}(t), C_{\text{Visibilty}}(t), \\ &C_{\text{WindSpeed}}(t), C_{\text{DewPoint}}(t), C_{\text{WindDirection}}(t)) \end{aligned} \tag{4}$$

The mean square error calculated during the model training and testing phase depends on the usage pattern of the appliances, the applied mode, and the input data. This error shall be employed as the fitness function during the hybrid optimization of PSO and ABC. Equation (5) was used as the fitness/objective function.

$$\text{MSE} = \frac{\sum_{i=1}^{n} (\hat{\text{Actual}}_i) - \text{Predicted}_i^2}{n} \tag{5}$$

3.3.2 *Training of neural network by PSO*

Particle swarm optimization was inspired from the swarm intelligence to find the shortest route for their activities (Eberhart and Kennedy 1995). It consists of particle update (p_i), velocity update (v_i) and comparing of their better values. The hybrid approaches used in this paper is to optimize the input parameters as described in Eq. (4). The input data was provided by the network created for ANN and the initialization of the weights were provided to the initial parameters of the PSO. The second step in PSO algorithm was to update the velocity and the position vector. Equations (6) and (7) describe the velocity update and position update.

$$\begin{aligned} V_{id} &= w \times v_{id} + const_1 \times ran_1 \times (pbest_{id} - x_{id}) \\ &\quad + const_2 \times ran_2 \times (gbest_{id} - x_{id}) \end{aligned} \tag{6}$$

where w is the interia weight, d is the number of parameters to be optimized, $const_1$ and $const_2$ are the acceleration constants and run_1 and run_2 are the random numbers between the range 0–1. The $const_1$ and $const_2$ are used to find the optimal path by moving the each particle towards the *pbest* and *gbest*.

$$X_{id} = x_{id} + V_{id} \tag{7}$$

The proposed hybrid algorithm ANN-PSO is described in Algorithm 1. The evaluation of the fitness function/ objective function used in Eq. (5) was calculated and the

Algorithm 1 ANN-PSO algorithm

```
1: Initialize the ANN network with input data
2: for
3:   all particles do
4:   Initialize the parameters of PSO
5: end for
6: loop
7:   for   do
8:     all particles
9:     Calculate the new velocity using Eq. (6)
10:    Calculate the new position using Eq. (7)
11:    Calculate the fitness value ate new position
12:   end for
13: Find the pbest value and set gbest value
14: end loop
15: If the termination condition is not satisfied go to Step 2
16: Bring the pbest and gbest value found in step 12 to the neural network
        initialize in Step 1
17: loop
18:   for do
19:     Calculate the optimal weight of the input data
20:   end for
21:   Calculate the MSE using Eq. (5)
22: end loop
23: If the termination condition is not satisfied go to Step 17
```

values of *pbest* and *gbest* were updated until the termination conditions satisfied. If the termination conditions are satisfied then the loop is passed to the neural network which is created at the initial phase for calculating the optimal weight and the mean square error by Eq. (5). Finally, the weights are updated and the model is ready to use for testing if the termination conditions are met otherwise the step is repeated until the conditions are not satisfied.

3.3.3 Training of neural network by ABC

Artificial bee colony (ABC) algorithm was inspired from the honey bees nature. It is a metaheuristic approach proposed by Karaboga (2005) and further developed by Karaboga and Akay (2007). It consists of three phases i.e. employed bee phase, onlooker bee phase and scout bee phase. The main goal is to produce the suitable solution for each individual bee. The optimization approach has been applied with artificial neural network to minimize the error rate (by using the Eq. (5)) and satisfied the minimum electricity demands of the dwellings for accomplishing their daily activities. The initial step was to train the neural network by initializing the value of the weights. The network thus created was provided along with the initialization parameters of the artificial bee colony algorithm. The employed bee located a new food source f_i in the region of the current source y_i. Equation (8) is used by the employed bee for locating the food source:

$$f_{jk} = y_{jk} + \Phi_{jk}(y_{jk} - y_{ik}) \tag{8}$$

where $i \in (1, 2, ..., SN)$ and $k \in (1, 2, ..., D)$ are the indexes which are randomly chosen and j and i need to be different from each other. Φ_{jk} is between −1 and 1 where j and k need to be different from each other. The greedy selection mechanism is used by the employed bee's for memorizing the better solution.

The second phase in ABC is known as onlooker bee phase that chooses the food source by using the probability according to the fitness function/objective function. Equation (9) is used for calculating the probability.

$$Prob_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_i} \tag{9}$$

The *fit* refers to the fitness function in the Eq. (5). The third, i.e., the final phase, is scout bee phase in which the food source cannot be improved by number of different cycles from the entire population and the employed bee of that source becomes scout. The new random source position f_i is found by scout bee using Eq. (10).

$$y_j^k = y_{\min}^k + \mathrm{ran}[0,1](y_{\max}^k - y_{\min}^k) \tag{10}$$

$y_{\min}^k$ and $y_{\max}^k$ are the upper and lower bounds of parameter k.

The three phases of the algorithm are repeated in cycles known as maximum cycle number (MCN) until a termination condition is satisfied. If a termination conditions are met the algorithm passes the value to the ANN network created in the initial phase of the algorithm for calculating the optimal weights and for mean square error. If the updated weight satisfies the condition the model is fully trained and ready of the testing otherwise the loop goes on till the condition is not satisfied. All the three phases explained above were implemented by using Algorithm 2.

Algorithm 2 ANN-ABC algorithm

```
Initialize the ANN network with input data
cycle=1
Initialize the food source position f_i
Fitness function evaluation of food source using Eq. (5)
Evaluate the source food and the value in gbest
repeat
  for do
    Each component y
    Employed Bee Phase
    for do
      Each employed bee i
      y component of gbest is replaced by using the y component
          of bee i
      Calculate the value of pbest
      if pbest > gbest then
        gbest is replaced by pbest
      endif
      New food source position is evaluated using Eq. (8)
      Fitness function evaluation of food source using Eq. (5)
    endfor
    Calculate the probability P_i using Eq. (9)
    Onlooker Bee Phase
     for do
      Each onlooker bee i
      Select f_i depending on P_i
      y component of gbest is replaced by using the y component
          of bee i
      Calculate the value of pbest
      if pbest > gbest then
        gbest is replaced by pbest
      endif
      New food source f_i position is evaluated
      Fitness function evaluation of food source using Eq. (5)
     endfor
  endfor
  Scout Bee Phase
  if employed bee = Scout then
      Replace it with new random source position
  endif
  Best solution is stored in some value
  Compare the solution with pbest
  Store the value
  cycle =cycle +1
until cycle = MCN
If the termination condition is not satisfied go to Step 2
loop
  for do
     Calculate the optimal weight of the input data
  endfor
  Calculate the MSE using Eq. (5)
endloop
If the termination condition is not satisfied go to Step 44
```

3.4 Evaluation parameters

There are different evaluation parameters on which the energy prediction models are compared. They are as follows:

– **Root Mean Square Error (RMSE):** It is the difference between the actual and the predicted value to the total number of inputs. The model which has less RMSE value is more fit for the energy prediction rather than the model having high RMSE value.

$$\mathrm{RMSE} = \sqrt{\frac{\sum (Predicted_i - Actual_i)^2}{n}} \tag{11}$$

where, $i = 1,2,3, \ldots, n$ and n is the total number of predictors.

– **Accuracy:** Accuracy is the measure of the percentage of the predicted and the actual values. It indicates "how accurately our model has performed the testing of the input data". It is calculated by Eq. (12).

$$\mathrm{Accuracy} = \frac{\sum_{i=1}^{n} (\mathrm{abs}(Actual_i - Predicted_i) \times 100 \le err}{n} \tag{12}$$

where, $Actual_i$ is the actual target, $Predicted_i$ is the predicted target, err is the acceptable error and n is the total number of instances.

Table 2 Appliance dataset (Makonin et al. 2013)

ID	Appliances
B1E	North Bedroom
B2E	Master/South Bedroom
BME	Basement Plugs & Lights
CDE	Clothes Dryer
CWE	Clothes Washer
DNE	Dining Room Plugs
DWE	Dishwasher
EBE	Electronics Workbench
EQE	Security/Network
FGE	Kitchen Fridge
FRE	HVAC/Furnace
GRE	Garage
HPE	Heat Pump
HTE	Instant Hot Water Unit
OFE	Home Office
OUE	Outside Plug
TVE	Ent Tv/PVR/AMP
UTE	Utility Room Plug
WOE	Wall Oven
RSE	Rental Home

Table 3 Clustered appliances

Cluster	Appliances
Cluster 1	UTE, EBE, OFE, EQE
Cluster 2	B1E,UNE, HTE, DWE, B2E, RSE, BME, TVE, DNE,GRE
Cluster 3	FRE, HPE, CDE, CWE, OUE

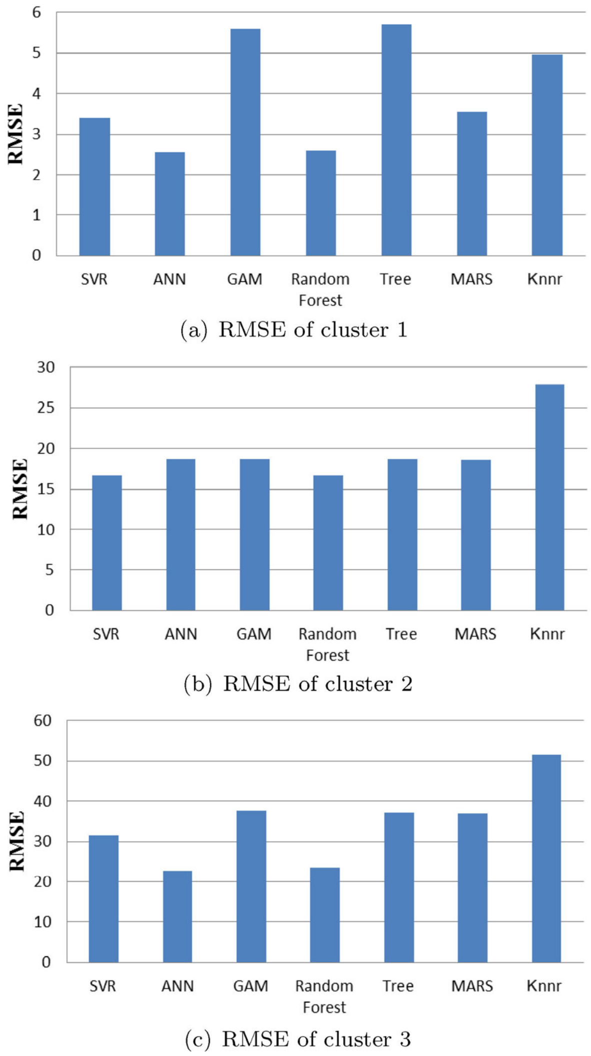
4 Experimental results

Principal component analysis was performed on the appliance data for reducing the dimensionality according to the usage pattern of appliances, and the K-means clustering was done. The three clusters were formed according to the usage pattern, all of which were integrated with the climate data-set. All the prediction experiments were conducted in R on the three clusters and the result was evaluated based on the accuracy and root mean square error. The most suitable prediction model was used with the optimization technique which resulted in the hybrid optimization. The two techniques of Particle Swarm Optimization and Artificial Bee Colony were performed in MATLAB 2016b. The experimental section was divided into two cases. Table 2 indicates the original appliance data-set (Makonin et al. 2013) that was clustered by using PCA, and the clustered appliances were merged with the climatic conditions. The cluster 1 appliances were UTE, EBE, OFE, and EQE; cluster 2 consisted of B1E, UNE, HTE, DWE, B2E, RSE, BME, TVE, DNE, and GRE; and lastly, cluster 3 consisted of FRE, HPE, CDE, CWE, and OUE as shown in Table 3.

CASE 1: Prediction models

The prediction models explained in Section 3.2 were implemented in R, and the different results for accuracy and root mean square were obtained by dividing the data-set into 70% training and 30% testing. All the regression machine learning algorithms explained above were executed. Figure 2 describes the root mean square error of all the models in all the three clusters formed during PCA. Figure 2(a) shows the results for cluster 1; ANN has the least RMSE value of 2.55 while the tree model has 5.59, which is the maximum value when compared with the other models. The RMSE of cluster 2, as seen in Fig. 2(b), is 18.65 for ANN, while the maximum value for Knnr model is 27.89. The ANN has the least value in cluster 2 too. The minimum RMSE in cluster 3 is for ANN model (31.58), while the maximum value is for the Knnr model (51.58), as inferred from Fig. 2(c). It could be understood that ANN achieved the least RMSE values in all the clusters.

Figure 3 depicts the accuracy results of all the three clusters. In cluster 1, the accuracy of ANN was maximum (99.96 %); however the accuracy values of all other models differed only slightly. The other models of SVR, and GAMs in cluster 1 achieved the accuracies of 99.49%, and 98.46%, respectively. The least accuracy of 78.34% was observed in

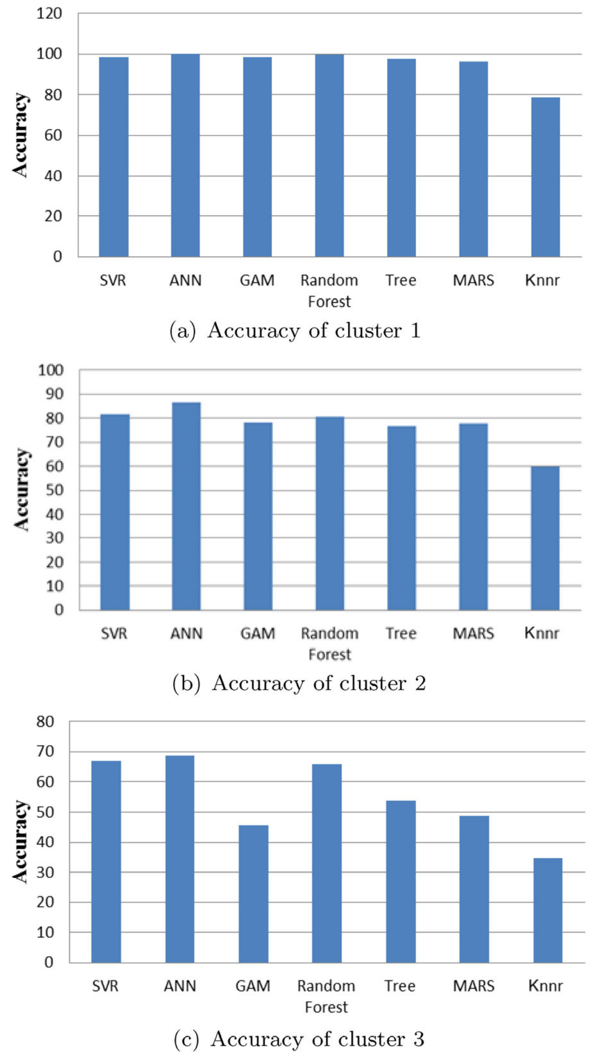


(a) RMSE of cluster 1

(b) RMSE of cluster 2

(c) RMSE of cluster 3

Fig. 2 RMSE values



(a) Accuracy of cluster 1

(b) Accuracy of cluster 2

(c) Accuracy of cluster 3

Fig. 3 Accuracy values

cluster 1 for Knnr. The accuracies of ANN in cluster 2 and cluster 3 were 86.37%, and 68.56%, respectively, while those achieved by the other models, namely SVR, random forest, and GAMs in cluster 2 were 81.76%, 80.62%, and 78.12%, respectively; in cluster 3, the accuracies were 66.88%, 65.95%, and 45.62%, respectively. When compared with the other models, the maximum accuracy from all the three clusters was obtained for ANN.

Table 4 provides the comparison between different evaluation parameters. The ANN performed better in energy prediction under different climatic conditions. The average accuracy attained by ANN (86.92%) was the maximum while that of Knnr (57.54%) was the minimum in terms of energy prediction for the household.

CASE 2: Optimization results of ANN-PSO

From the prediction results, it is clear that ANN has the minimum error value and the maximum accuracy for all the three clusters. Hence, this method was used along with the optimization technique to achieve the optimal energy values. The ANN-PSO was implemented for minimizing the energy consumption value. Figure 4 illustrates the comparison of actual, ANN model, and ANN-PSO predicted values. The cluster 1 is represented in Fig. 4(a), and it describes the average energy consumption values of actual (4.83), ANN model (8.09), and ANN-PSO (6.09) for one day. Thus, it can be clearly validated that the ANN-PSO and ANN model differ only slightly from the actual energy value. The average energy consumption values for cluster 2 are exhibited in Fig. 4(b). The actual (39.32), ANN model (38.40), and ANN-PSO (38.28) values for cluster 2 are denoted. Finally, the values of actual (117.19), ANN model (82.54), and ANN-PSO (115.34) for cluster 3 are presented in Fig. 4(c).

CASE 3: Optimization results of ANN-ABC

The ANN-ABC was implemented for minimizing the energy consumption value. Figure 5 depicts the values for all the clusters. The actual (4.83), ANN model (8.09), and ANN-ABC

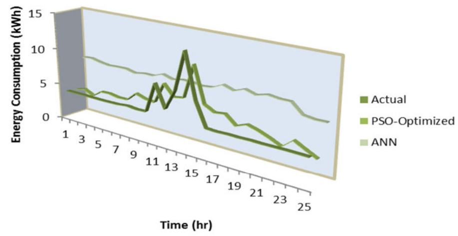
Table 4 Results of prediction models

Model name	Cluster No.	RMSE	Accuracy (%)
ANN	1	2.55	98.11%
	2	15.23	92.50%
	3	22.59	70.17%
	Average	13.45	86.92%
SVR	1	3.41	98.04%
	2	16.71	81.76%
	3	31.58	66.88%
	Average	17.23	78.52%
Random Forest	1	2.59	97.89%
	2	16.65	80.62%
	3	23.39	65.95%
	Average	14.21	81.48%
Tree	1	5.59	97.54%
	2	18.73	76.66%
	3	37.27	53.91%
	Average	20.53	76.03%
MARS	1	3.54	96.35%
	2	18.63	77.66%
	3	36.9	48.62%
	Average	19.69	74.21%
GAM	1	5.59	97.46%
	2	18.74	78.12%
	3	37.62	45.62%
	Average	20.65	73.73%
Knnr	1	4.96	78.34%
	2	27.89	59.69%
	3	51.58	34.6%
	Average	28.14	57.54%

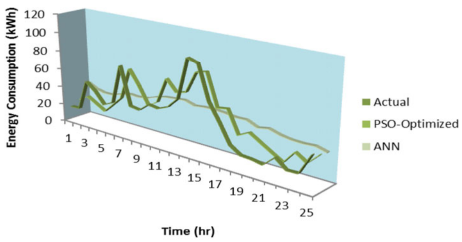
(5.05) values of cluster 1 are portrayed Fig. 5(a). The actual (39.32), ANN model (38.40) and ANN-ABC (37.13) energy consumption values of cluster 2 are signified in Fig. 5(b). Therefore, it could be inferred that ANN-ABC performed better than the ANN model. The per day energy consumption values for cluster 3 are shown in Fig. 5(c).

CASE 4: Comparison of ANN, ANN-PSO, ANN-ABC

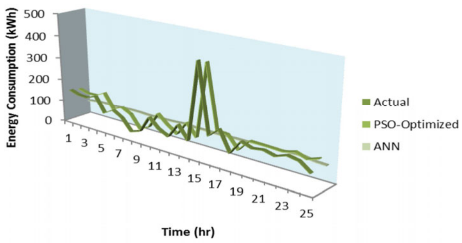
The comparison of the average value of energy consumption per day among the actual, ANN, ANN-PSO, and ANN-ABC is furnished in Fig. 6. The average actual (4.83), ANN model (8.09), ANN-PSO (6.09) and ANN-ABC (5.05) values of cluster 1 are presented in Fig. 6(a). Similarly, the energy one day values for clusters 2 and 3 are shown in Figs. 6(b) and (c). The average energy value obtained for ANN-PSO was (6.09) and for ANN-ABC (5.05) for cluster 1. Therefore, it can be clearly deducted that the ANN-ABC performed better than ANN-PSO.



(a) ANN-PSO result of cluster 1

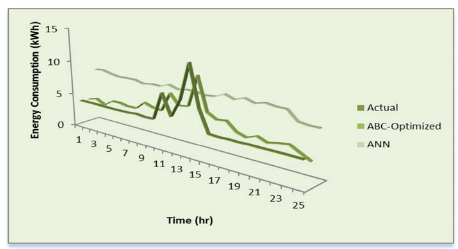


(b) ANN-PSO result of cluster 2

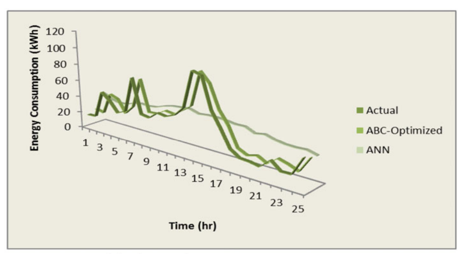


(c) ANN-PSO result of cluster 3

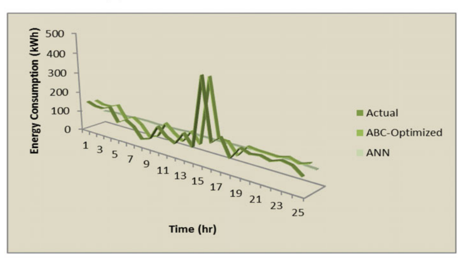
Fig. 4 ANN-PSO optimized results



(a) ANN-ABC result of cluster 1

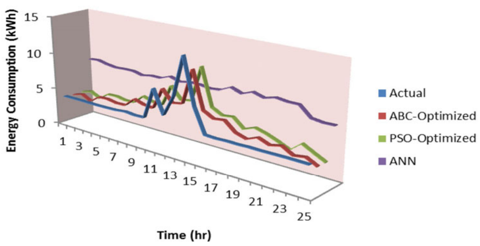


(b) ANN-ABC result of cluster 2

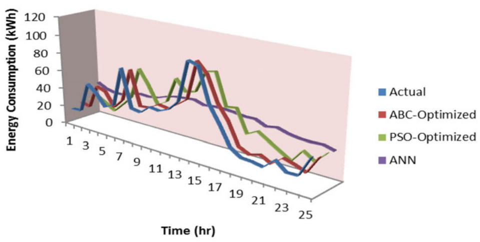


(c) ANN-ABC result of cluster 3

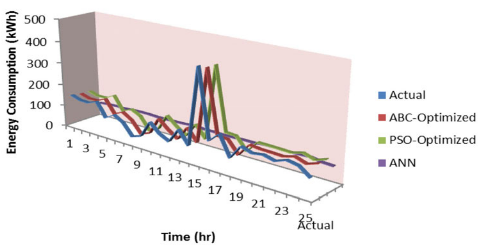
Fig. 5 ANN-ABC optimized results



(a) Comparison of ANN , ANN-PSO, ANN-ABC of cluster 1



(b) Comparison of ANN , ANN-PSO, ANN-ABC of cluster 2



(c) Comparison of ANN , ANN-PSO, ANN-ABC of cluster 3

Fig. 6 Comparison of ANN, ANN-PSO, ANN-ABC results

Table 5 Optimized energy consumption (kWh)

Cluster No.	Actual energy consumption for one day	Predicted energy consumption for one day	ANN-PSO	ANN-ABC
1	8.09	7.15	5.09	5.00
2	41.12	39.32	38.28	37.81
3	117.34	115.34	114.94	82.53
Average	166.55	161.81	158.31	125.43

Table 5 provides details of the per-day energy consumption in a household. The actual average (166.55 kWh), predicted (161.81 kWh), ANN-PSO optimized (158.31 kWh) and ANN-ABC optimized (158.31 kWh) values for a households energy consumption are given. From the table, it can be clearly concluded that the energy consumption can be minimized by using ANN-ABC.

From cases 2, 3, and 4, it is clear that the average energy value for one day obtained by the ANN model is higher than those obtained by the other optimization models for all the three clusters. Hence, the ANN-ABC optimization performed better than the ANN-PSO optimization in all the three clusters. The techniques can thus be employed by the energy suppliers to meet the minimum electricity requirement of the dwellings.

5 Comparative analysis

As per the literature review (Gajowniczek and Ząbkowski 2017), a comparison of some of the machine learning models with the existing models was done using the same dataset.

The optimization of energy was not performed. Only the prediction was carried out, which was not based on the cluster formation but in the proposed work, the prediction was based on the clusters along with the climatic conditions. The average values of all the clusters were compared with the existing model. Figure 7 compares the accuracy results of the existing model with the proposed model. The ANN in the existing model achieved a very less accuracy of 42.26%; in the proposed model, it achieved the maximum average accuracy of 86.92%. The random forest model achieved the maximum accuracy of 100% among the existing models; in the proposed model, it achieved an average accuracy of 81.48% because the climatic conditions were not considered in the existing models. The other existing models, such as SVR (47.02%), Tree (35.12%), and Knnr (41.37%), were less accurate than the proposed models of SVR (82.22%), Tree (76.03%), and Knnr (41.37%).

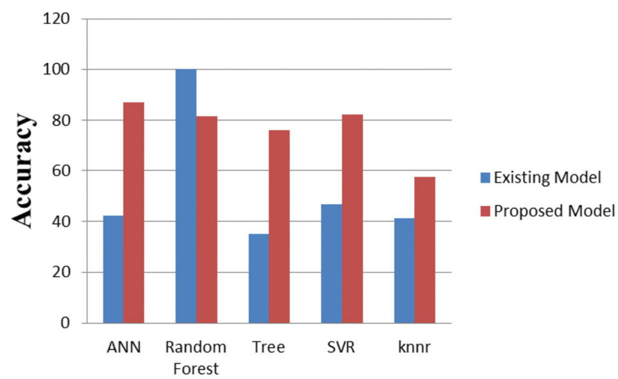


Fig. 7 Comparison of proposed model with existing model

6 Conclusion

In this paper, the climatic conditions were considered for predicting the energy demand in the homes according to the previous pattern of consumption. The consumption pattern of the appliance usage was determined by the principal component analysis and the clustering was performed based on it. The different clusters were further merged with the climatic conditions. The machine learning algorithms were implemented for predicting the energy demand and the model yielding superior results was used along with the optimization technique. The average accuracy of ANN (86.92%) was the highest when compared with the other models and that of Knnr (57.54%) was the least for forecasting the energy requirement of the household. The ANN model achieved the accuracies of 99.96%, 86.37%, and 68.56%

in clusters 1, 2, and 3, respectively. Hence, this model was merged particle swarm optimization and artificial bee colony. The ANN-PSO and ANN-ABC were used as the optimization techniques. The average per-day values of the actual (4.83), ANN model (8.19), ANN-PSO (6.09), and ANN-ABC (5.09) for cluster 1 were recorded. Similarly, the actual (39.32), ANN model (38.40), ANN-PSO (38.28), and ANN-ABC (37.13) values for cluster2 were documented. Likewise, for cluster 3, the actual (117.19), ANN model (82.54), ANN-PSO (115.34), and ANN-ABC (114.94) values were noted. Upon comparison, it was observed that the ANN-PSO and ANN-ABC performed better than the ANN model. The optimization results enabled usto conclude that the energy suppliers need to provide the minimum amount of energy to the dwellings as per the aforementioned values so that; they could accomplish their basic daily activities. Further, energy forecasting can be done by taking the different factors that influence the energy usage in homes, such as number of occupants, demographic factor, and behaviour of the occupant.

In future, a feedback mechanism can be made available for the occupants about the supply of electricity in their dwellings. The energy suppliers can build a system to store the extra energy for the future use of the occupants. The method can be used for prediction of the energy requirement in buildings and can be expanded to other sectors, such as industrial, agriculture, and transportation, which are facing an escalation in the energy demand. The system can be scaled-up by using deep learning and artificial intelligence methods. It can further be utilized to optimize the energy for electrical smart-grid to realize real life applications.

References

Ahmad AS, Hassan MY, Abdullah MP, Rahman HA, Hussin F, Abdullah H, Saidur R (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33: 102–109.

Ardakani FJ, Ardehali MM (2014). Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. *Energy*, 65: 452–461.

Balshi MS, McGuire AD, Duffy P, Flannigan M, Walsh J, Melillo J (2009). Assessing the response of area burned to changing climate in Western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology*, 15: 578–600.

Biswas MAR, Robinson MD, Fumo N (2016). Prediction of residential building energy consumption: A neural network approach. *Energy*, 117: 84–92.

Candanedo LM, Feldheim V, Deramaix D (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140: 81–97.

Chen Y, Xu P, Chu Y, Li W, Wu Y, Ni L, Bao Y, Wang K (2017). Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Applied Energy*, 195: 659–670.

Cottone P, Gaglio S, Re GL, Ortolani M (2015). User activity recognition for energy saving in smart homes. *Pervasive and Mobile Computing*, 16: 156–170.

Divina F, Gilson A, Goméz-Vela F, García Torres M, Torres J (2018). Stacking ensemble learning for short-term electricity consumption forecasting. *Energies*, 11(4): 949.

Do H, Cetin KS (2018). Residential building energy consumption: a review of energy data availability, characteristics, and energy performance prediction methods. *Renewable Energy Reports*, 5: 76–85.

Dong B, Li Z, Rahman SM, Vega R (2016). A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, 117: 341–351.

Eberhart R, Kennedy J (1995). A new optimizer using particle swarm theory. In: Proceedings of the 6th *International Symposium on Micro Machine and Human Science (MHS'95), Nagoya, Japan,* pp. 39–43.

Gajowniczek K, Ząbkowski T (2017). Electricity forecasting on the individual household level enhanced based on activity patterns. *PLoS One*, 12: e0174098.

Godina R, Rodrigues E, Pouresmaeil E, Matias J, Catalão J (2016). Model predictive control technique for energy optimization in residential sector. In: Proceedings of the 16th International Conference on Environment and Electrical Engineering (EEEIC), Florence, Italy.

Gupta N, Ahuja N, Malhotra S, Bala A, Kaur G (2017). Intelligent heart disease prediction in cloud environment through ensembling. *Expert Systems*, *34* (3): e12207.

Ha DL, Joumaa H, Ploix S, Jacomino M (2012). An optimal approach for electrical management problem in dwellings. *Energy and Buildings*, 45: 1–14.

Juan Y-K, Gao P, Wang J (2010). A hybrid decision support system for sustainable office building renovation and energy performance improvement. *Energy and Buildings,* 42: 290–297.

Karaboga D (2005). An idea based on honey bee swarm for numerical optimization. Technical report-tr06. Erciyes University, Turkey.

Karaboga D, Akay B (2007). Artificial Bee Colony (ABC) algorithm on training artificial neural networks. In: Proceedings of IEEE 15th Signal Processing and Communications Applications (SIU), Eskisehir, Turkey.

Kaur J, Bala A (2018). Predicting power for home appliances based on climatic conditions. *International Journal of Energy Sector Management*, https://doi.org/10.1108/IJESM-04-2018-0012

Kaur J, Bala A (2019). Review of machine learning techniques for optimizing energy of home appliances. In: Fong S, Akashe S, Mahalle PN (eds), Information and Communication Technology for Competitive Strategies. Singapore: Springer Singapore. pp. 255–263.

Khakimova A, Kusatayeva A, Shamshimova A, Sharipova D, Bemporad A, Familiant Y, Shintemirov A, Ten V, Rubagotti M (2017). Optimal energy management of a small-size building via hybrid model predictive control. *Energy and Buildings*, 140: 1–8.

Lazos D, Sproul AB, Kay M (2014). Optimisation of energy management in commercial buildings with weather forecasting inputs: A review. *Renewable and Sustainable Energy Reviews*, 39: 587–603.

Li K, Pan L, Xue W, Jiang H, Mao H (2017). Multi-objective optimization for energy performance improvement of residential buildings: A comparative study. *Energies*, 10(2): 245.

Lin S-W, Ying K-C, Chen S-C, Lee Z-J (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35: 1817–1824.

Mahmood D, Javaid N, Alrajeh N, Khan Z, Qasim U, Ahmed I, Ilahi M (2016). Realistic scheduling mechanism for smart homes. *Energies*, 9(3): 202.

Makonin S, Popowich F, Bartram L, Gill B, Bajić IV (2013). AMPds: A public dataset for load disaggregation and eco-feedback research. In: Proceedings of IEEE Electrical Power & Energy Conference, Halifax, Canada.

Makonin S, Ellert B, Bajić IV, Popowich F (2016). Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific Data*, 3: 160037.

Muralitharan K, Sakthivel R, Vishnuvarthan R (2018). Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing*, 273: 199–208.

Rasheed M, Javaid N, Ahmad A, Jamil M, Khan Z, Qasim U, Alrajeh N (2016). Energy optimization in smart homes using customer preference and dynamic pricing. *Energies*, 9(8): 593.

Sharma N, Sharma P, Irwin D, Shenoy P (2011). Predicting solar generation from weather forecasts using machine learning. In: Proceedings of IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium.

Singh S, Yassine A (2017). Mining energy consumption behavior patterns for households in smart grid. *IEEE Transactions on Emerging Topics in Computing*, https://doi.org/10.1109/TETC. 2017.2692098

Subbiah R, Pal A, Nordberg EK, Marathe A, Marathe MV (2017). Energy demand model for residential sector: A first principles approach. *IEEE Transactions on Sustainable Energy*, 8: 1215–1224.

Swan LG, Ugursal VI (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13: 1819–1835.

Tian W, Song J, Li Z (2014). Spatial regression analysis of domestic energy in urban areas. *Energy*, 76: 629–640.

Torabi M, Hashemi S, Saybani MR, Shamshirband S, Mosavi A (2019). A hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental Progress & Sustainable Energy*, 38: 66–76.

Wang X, Palazoglu A, El-Farra N (2015). Operational optimization and demand response of hybrid renewable energy systems. *Applied Energy*, 143: 324–335.

Wang Z, Wang Y, Srinivasan RS (2018a). A novel ensemble learning approach to support building energy use prediction. *Energy and Buildings*, 159: 109–122.

Wang Z, Wang Y, Zeng R, Srinivasan RS, Ahrentzen S (2018b). Random forest based hourly building energy prediction. *Energy and Buildings*, 171: 11–25.

Wei Y, Zhang X, Shi Y, Xia L, Pan S, Wu J, Han M, Zhao X (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82: 1027–1047.

Wood SN (2001). mgcv: GAMS and generalized ridge regression for R. *R News*, 1(2): 20–25.

Yin P-Y, Chao C-H (2018). Automatic selection of fittest energy demand predictors based on cyber swarm optimization and reinforcement learning. *Applied Soft Computing*, 71: 152–164.

Zhu G, Chow T-T, Tse N (2018). Short-term load forecasting coupled with weather profile generation methodology. *Building Services Engineering Research and Technology*, 39: 310–327.