
# Ensemble learning-based approach for residential building heating energy prediction and optimization

Zhang Jianxin, Huang Yao, Cheng Hengda, Chen Huanxin, Xing Lu, He Yuxuan

PII: S2352-7102(23)00230-9

DOI: https://doi.org/10.1016/j.jobe.2023.106051

Reference: JOBE 106051

To appear in: *Journal of Building Engineering*

Received Date: 8 September 2022

Revised Date: 30 December 2022

Accepted Date: 1 February 2023

Please cite this article as: Z. Jianxin, H. Yao, C. Hengda, C. Huanxin, X. Lu, H. Yuxuan, Ensemble learning-based approach for residential building heating energy prediction and optimization, *Journal of Building Engineering* (2023), doi: https://doi.org/10.1016/j.jobe.2023.106051.

# Ensemble Learning-based Approach for Residential Building Heating Energy Prediction and Optimization

Zhang Jianxin[1] Huang Yao[2] Cheng Hengda[2] Chen Huanxin[2,*] Xing Lu[3] He Yuxuan[2]

[1]China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China;

[2]Department of Refrigeration & Cryogenics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[3]Mechanical and Construction Engineering, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom

**Abstract** Accurate building energy consumption prediction is critical for engineers to design optimized operational strategies for building heating, ventilation, and air-conditioning systems. In this paper, an stacking ensemble learning-based model is established based on the operational data of a district resident buildings heating station for building heating system energy consumption prediction. The ensemble model is optimized by outlier processing, feature selection, parameter optimization based on grid search. A new feature based on Exponentially Weighted Moving Average (EWMA) algorithm was proposed to take historical energy feature into consideration. The performance of the ensemble model and four base machine learning methods, including multiple linear regression, extreme learning machine, extreme gradient boosting and support vector regression, are evaluated. Compared with the four base models, the Mean Absolute Error (MAE) of the ensemble model decreases by 4.36% ~ 71.70%, and the Root Mean Squared Error (RMSE) by 3.80% ~ 49.73%. Using the new feature based on EWMA can further reduce the MAE and RMSE of the ensemble model by

10.36% and 19.89%, respectively. The result proves that the proposed ensemble model with the added historical feature effectively improves the prediction model's accuracy for building heating energy consumption.

**Keywords** Energy consumption prediction; Ensemble learning; Heating Station; Machine Learning; Optimization

| **Nomenclature** | **Unit** | **Description** |
|---|---|---|
| HVAC | / | Heating, Ventilation, and Air Conditioning |
| EWMA | / | Exponentially weighted moving average |
| DWT | / | Discrete wavelet transform |
| MIC | / | Maximal information coefficient |
| ELM | / | Extreme learning machine |
| GBDT | / | Gradient boosting decision tree |
| BPNN | / | Back propagation neural network |
| MLR | / | Multiple linear regression |
| PCA | / | Principal component analysis |
| SVM | / | Support vector machine |
| SVR | / | Support vector regression |
| XGB | | Extreme gradient boosting |
| $T_{sup}$ | ℃ | Water supply temperature |
| $T_{back}$ | ℃ | Return water temperature |
| $P_{sup}$ | kPa | Water supply pressure |

| | | |
|---|---|---|
| $P_{back}$ | kPa | Return water pressure |
| $m$ | $m^3/h$ | Instantaneous flow rate |
| EC | MJ | Instantaneous heat |
| $T_{out}$ | ℃ | Outdoor temperature |
| $T_{dew}$ | ℃ | Dew point temperature |
| RH | % | Relative humidity |
| $V_w$ | m/s | Outdoor wind speed |
| $W$c | / | Weather conditions |
| MAE | / | Mean absolute error |
| RMSE | / | Root mean squared error |

## 1. Introduction

Buildings' energy consumption prediction is an initial and critical step to tackling building energy-saving problems, which enables engineers to optimize the operational strategy of the mechanical Heating, Ventilation, and Air Conditioning (HVAC) system. HVAC systems consume much energy and predicting the HVAC system energy consumption is essential. The HVAC energy consumption is affected by many factors, such as climate, materials and structure, human behavior, and the HVAC operational status[1]. In the past decades, researchers proposed various forecasting methods to improve the accuracy of energy consumption prediction, most of which are based on historical data on HVAC energy consumption. There are three main HVAC energy consumption forecast models: forward modeling, statistical, and data-driven.

The forward modeling method builds a physical model of the whole building or building subsystem and simulates the HVAC energy consumption process of the system based on thermodynamics and heat transfer principles. Shabunko et al. [2] predicted the HVAC energy consumption of more than 400 residential buildings by simulation and divided more than 400 residential buildings into three types. The building's annual energy use intensity is calculated based simulation model, and energy building simulation models are verified against utility data and surveys. The results show that the simulated data can demonstrate the capability of engineering models. Bansal et al. introduced a room temperature prediction model based on TRNSYS. The room used a new phase change material. The model was verified by experiment data and actual building monitoring data. The results show that the model established by TRNSYS has

good consistency with actual building data and can be applied to the temperature control system[3].

Statistical theory and methods have been widely used in building energy consumption prediction before using the Multiple Linear Regression (MLR) model to predict building HVAC energy consumption. MLR has the advantages of simple structure and no need for parameter adjustment, but it has the disadvantages of low prediction accuracy and long computational time[4]. Besides MLR, Principal Component Analysis (PCA) is widely used in energy consumption prediction models[5]. Li et al. chose to build an energy consumption prediction model that uses the variable set obtained by PCA as input. The prediction results show that using the variable set obtained by PCA can effectively improve the model's prediction accuracy and shorten the computational time [6]. Gong et al. proposed an accurate prediction model for residential buildings' short-term heat load consumption based on discrete wavelet transform (DWT) and highly randomized tree regression. The results show that by applying DWT and the feature selection based on the least absolute shrinkage and selection operator method, the extreme random tree model achieved better prediction performance and lower model complexity[7].

With the rapid development of machine learning, more and more scholars have applied machine learning algorithms to predict building energy consumption. A traditional method such as Back Propagation Neural Network (BPNN) and Support Vector Machines (SVM) are often used in the research, and ensemble models are also popular methods and can provide state-of-art results[8, 9]. Li et.al.[10] established a

prediction model of the hourly cooling load of buildings based on SVM and applies it to an office building in Guangzhou. Compared with the BPNN model, the SVM model has better accuracy and is more robust in predicting the building cooling load. Golkarnarenji et al. [11] predicted energy consumption in the chemical industry by SVM, and optimized penalty factor C and range factor $\gamma$ by Genetic Algorithm, which have impact on the general ability of SVM. Thus, an intelligent energy consumption prediction model based on production quality and control of stochastic defects in thermal stability process was established. The energy consumption was reduced by 40% based on the predicted data. Wang Zeyu et.al. [12] proposed a model for predicting the hourly electricity demand of buildings based on ensemble bagging trees and data obtained from meteorological systems and building-level occupancy and meters. The result showed that the MAE of the proposed ensemble bagging trees model in predicting hourly electricity demand of the test building ranges from 2.97% to 4.63%, and the computation time was reduced. Eric Ofori-Ntow Jnr et.al. [13] proposed a three-level hybrid ensemble short-term load forecasting method consisting of DWT, Particle Swarm Optimization, and Radial Basis Function Neural Network, which achieved good performance in predicting the daily electricity demand of Ghana Grid Company. Wang Xuan et.al. [14] proposed a multi-energy load prediction model based on deep multi-task learning and ensemble approach for a regionally integrated energy system based on a hybrid network of convolutional neural network and gated recurrent unit with homoscedastic uncertainty and used ensemble approach based on gradient boosting regressor tree (GBRT) to make a weighted summary by the prediction

results of various energy features learning in different degrees.

This paper proposed an ensemble learning-based model for predicting residential buildings' HVAC heating energy consumption. The ensemble model utilized stacking strategy to combine extreme gradient boosting, extreme learning machine, and multiple linear regression as first-level models and set support vector regression as second-level model. The ensemble model is developed based on historical energy data of resident buildings' heating stations. In order to improve the stability and reliability of the model, a feature selection process based on a random forest algorithm was applied, and a comprehensive feature based on exponentially weighted moving average was added. The proposed ensemble model performance is evaluated through the test set, and the result is compared with four models previously developed by other researchers. The ensemble model's accuracy and reliability have been analyzed.

The organization of this paper is as follows: Section 2 outlines the methodology of the proposed energy consumption prediction model. Section 3 provides the introduction and input of the data used for establishing the model evaluation of the proposed models. Section 4 describes the proposed model's results and its discussion. Section 5 describes the conclusion of the study.

## 2. Methods

The framework of the proposed method is illustrated in fig.1. The process includes data preprocessing, model training, model optimization, and evaluation.

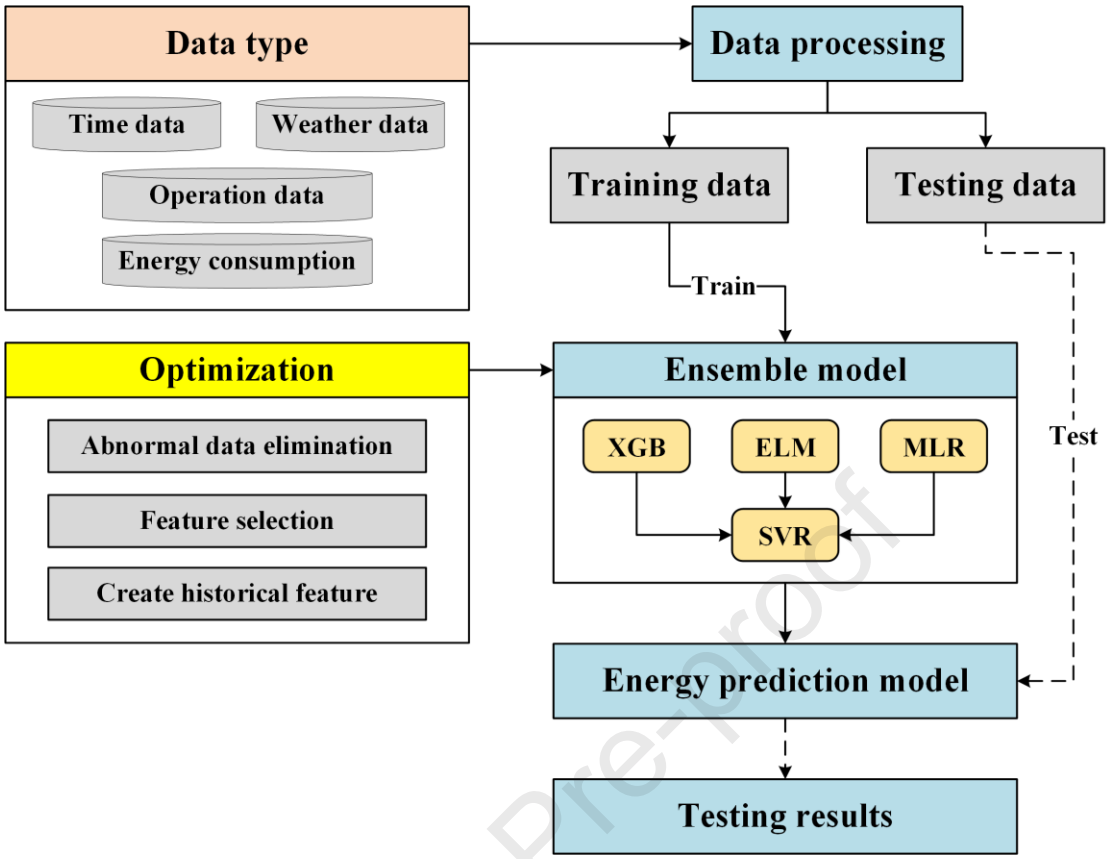



Fig.1 Flow chart of proposed ensemble learning-based model

## 2.1 Ensemble learning method

The essence of ensemble learning is constructing many base machine learning models for learning and then combining multiple machine learning models through specific strategies, thus accomplishing a task and achieving better results than those base machine learning methods[15]. In the ensemble model, the data set will be divided and used to train different machine learning models. According to optimization and the dependence between individual learners, ensemble learning is generally divided into three categories[16]: bagging for reducing variance, boosting for reducing deviation, and stacking for improving prediction effect. Thus, the stacking strategy was utilized in this study. The general architecture of ensemble learning is shown in fig.2.

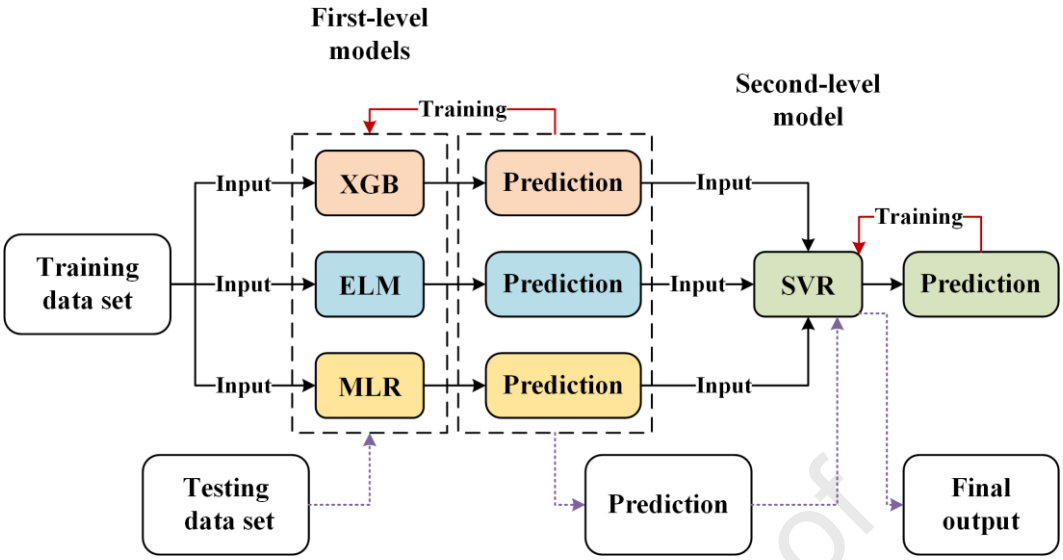



Fig.2 Ensemble learning schematic

In the ensemble model, the data set will be divided and used to train different machine learning models called base models. The extreme gradient boosting, extreme learning machine and multiple linear regression were adopted as base models in first level. The training data set will be used to train each base model and the predicted results of each base model were combined as new input data, which will be used to train the second-level model, known as meta model. The support vector regression was set as the second-level model to get final prediction. Selecting proper base model is the key to establishing an ensemble learning model with good prediction ability. The brief introduction of the base machine learning algorithm is as follows:

The multivariate linear regression(MLR) algorithm is a quantitative analysis method. A multivariate linear regression model is established by analyzing the relationship between a dependent variable and multiple independent variables based on historical sample data[17]; its equation can be expressed as follows, where $\hat{Y}$ is the dependent variables.

$$\hat{\Upsilon} = \sum_{i=1}^{k} \beta_i x_i + \beta_0 \quad (1)$$

The extreme learning machine algorithm (ELM) [18] is a method proposed by Huang Guangbin et al. The ELM algorithm here is a single hidden layer feedforward neural network with significant advantages in calculation speed, and a random gradient is used to prevent it from falling into the optimal local situation in training.

Gradient Boosting Decision Tree (GBDT) is a representative method among Extreme Gradient Boosting (XGB) [19]. In the GBDT method, a decision tree is trained on the original training set $\Psi_0$ and then the residual value set $\varepsilon_i$ The trained model can be calculated by subtracting the predicted and actual values. After that, new data combined $\Psi_0$ and $\varepsilon_i$ is constructed, and a second tree will be trained on the new data set with the residual $\varepsilon_i$ used as the "true value," and the residual of each sample will be calculated again. This process will be repeated until the number of trees or the residual value satisfies the termination condition. When a new sample is an input to GBDT, each tree will output a value, which will be added to form the final prediction value.

Support vector regression(SVR) is a crucial application branch of SVM that can be used as a regressor. SVM is a classifier that can non-linearly map the training data samples to the high dimensional feature space and search for the optimal hyperplane to maximize the edge space between different categories to classify samples. Similarly, SVR can also find a hyperplane according to its input. The hyperplane found by SVR is not used to distinguish two or more types of sample points like SVM but to minimize the total deviation of all sample points from the hyperplane and achieve regression.

## 2.2 Model training

This research divides the data set into training and test sets. The train set is used for training the model, while the test set is used to evaluate the model's accuracy. The partition of the train set and test set directly affects the accuracy of the final model. Cross-validation is a statistical analysis method used to verify the performance of classifiers[20]. It divides the data into several sets and uses 1 set as a test set and the rest as train sets, and the test sets are changed in turn to train and evaluate the model multiple times. A sample in the training set will become a sample in the test set next time. With cross-validation used, the model can make full use of the data. The final result in cross-validation is the average value of all evaluations. Different data sets are used for training and evaluated in cross-validation; it can also prove that the model has good generation ability if it achieves good accuracy in the result.

## 2.3 Model optimization

The prediction accuracy of the model is improved by constantly adjusting the parameters of the model. The grid search is an optimization method used to find the optimal parameters of the model. The grid search method will set all the possible values of each model parameter and list all possible combinations of parameters to generate a "grid." The model will be trained with all the combinations in the "grid" to find the optimized combination of parameters[21].

In this paper, grid search is used to optimize the ensemble prediction model with 10-fold cross-validation used, and the performance of each combination of parameters is evaluated. Thus the optimal parameter combinations of the ensemble prediction

model can be found.

## 2.4 Model result evaluations

Model evaluation is essential in establishing the model and helps identify the best model in the result. The evaluation criteria for the model are MAE (mean absolute error) and RMSE (Root Mean Square Error). The formula of MAE[22] and RMSE[23] can be described as follows:

$$MAE = \frac{\sum_{i=1}^{n} | y_i - \hat{y}_i |}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{3}$$

Where n is the total number of samples, $y_i$ is the actual energy consumption value and $\hat{y}_i$ is the predicted energy consumption value. The smaller MAE and RMSE values in the evaluation, the higher the accuracy of the ensemble model.

# 3. Case study

## 3.1 Building heating system

The resident buildings studied in this paper are located in Chaoyang District, Beijing, with a total floor area of 90,000 square meters. It includes seven apartments and one office building, containing over 800 rooms. This research studied the heating season of building systems from November 15 to March 15 of next year. In this research, the HVAC systems in residents' buildings operate all day during the heating season. District heating is served by the heating station shown in fig.3. The heating station can adjust the heat network's operation status, transport heat to the end users according to

their demand, and provide records of parameters during the system operation[24]. The leading equipment of the heating station includes a heat exchanger, circulating pump, decontaminator, water supply pump, steam trap, water tank, distribution equipment, and metering equipment.

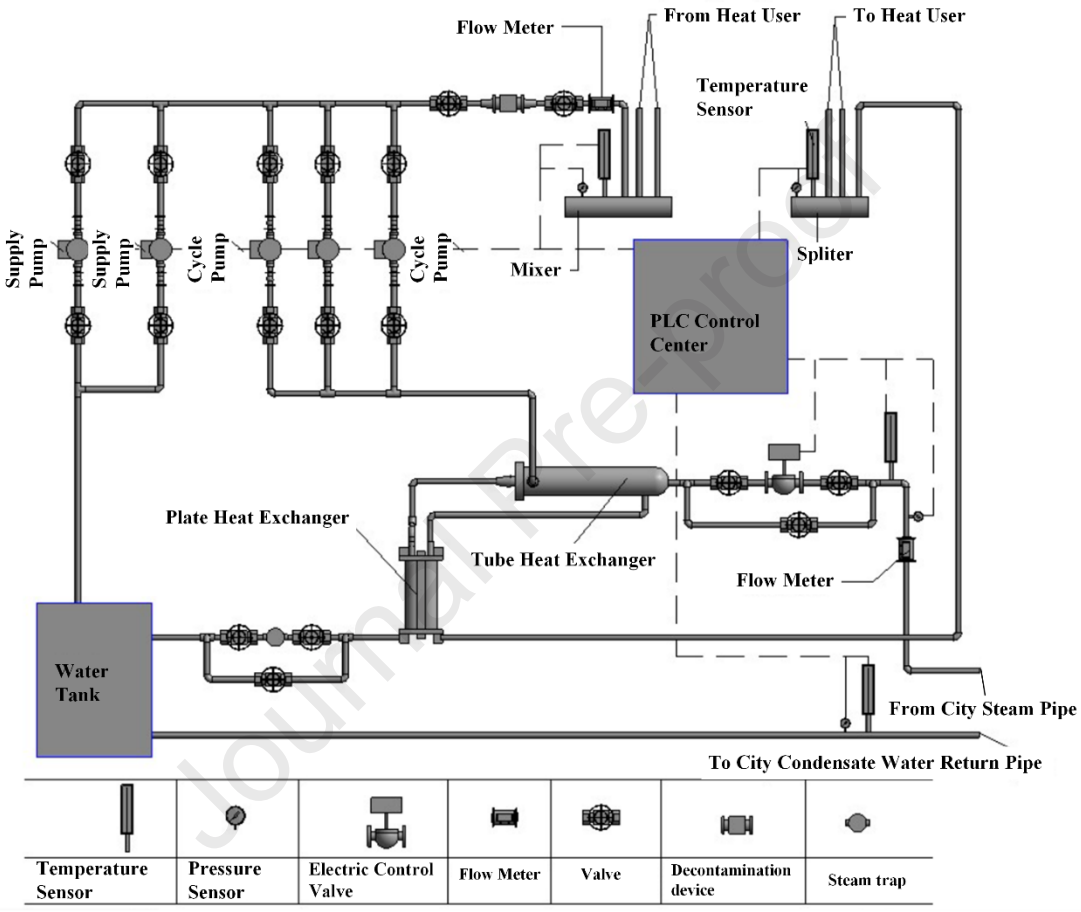



Fig.3 Heating station system structure

## 3.2 Model input

The data source used as input of the ensemble model includes data from the sampling period ranging from 2016/11/15 08:00 to 2017/3/15 08:00, covering the whole heating season. The water supply temperature ($T_{sup}$), return water temperature ($T_{back}$), water supply pressure ($P_{sup}$), return water pressure ($P_{back}$), instantaneous flow rate (m), and instantaneous heat (EC) in the secondary network of the high area in the operation

data is selected in this research. Meteorological parameters include outdoor dry bulb temperature ($T_{out}$), dew point temperature ($T_{dew}$ ), relative humidity (RH), outdoor wind speed ($V_w$), and weather conditions (Wc) are also selected. The sampling interval between operational data and meteorological parameters is different. In order to ensure the uniformity of data at each time, the sampling interval is set to 30 minutes. The rearranged data set includes 5760 samples, and all variables involved in the modeling and their ranges are listed in table 1. In order to achieve better accuracy in energy consumption prediction, the heating system of residential buildings in heating season data is further preprocessed by conducting abnormal value removal, feature selection, and adding a new variable, EMWA.

Table.1 Model parameters

| Variable category | Variable name | Variable range |
|---|---|---|
| **Meteorological parameters** | $T_{out}$ (℃) | -15 ~ 20 |
| | $T_{dew}$(℃) | -28~ 7 |
| | RH | 0.04 ~ 1.00 |
| | WC | Sunny, smog, Cloudy, etc. |
| | $V_w$(m/s) | 0.00 ~ 64.80 |
| **Operation parameters** | $T_{sup}$ (℃) | 29.50~65.40 |
| | $T_{back}$ (℃) | 29.30 ~ 60.00 |
| | $P_{sup}$ ($kPa$) | 0.840 ~1.080 |
| | $P_{back}$ ($kPa$) | 0.770 ~ 0.950 |

| | | |
|---|---|---|
| | $m$ ($m^3/h$) | 0 ~ 24.00 |
| | $EC$ ($MJ$) | -1.89 ~ 921.98 |

### 3.2.1 Abnormal value processing

As shown in table 1, evident abnormal values exist in some variables, such as negative instantaneous heat. If the original data set is used for modeling, the outlier may influence the model's prediction accuracy and even lead to modeling failure. Therefore, statistical and machine learning methods are used to detect outliers, analyze their causes, and then replace the outliers to improve the accuracy of the prediction model.

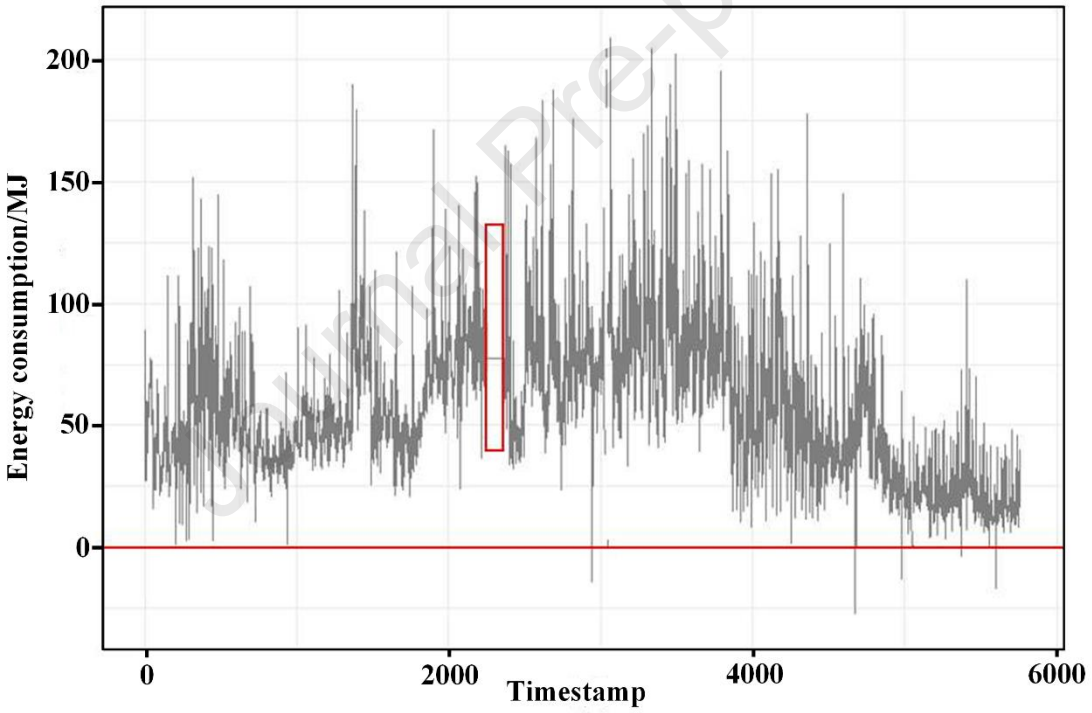



Fig.4 System instantaneous heat map

Fig.4 is the curve plot of instantaneous heat in the data set. It shows that the instantaneous heat value remained unchanged during 2017/1/1~2017/1/3. The flow rate and pressure also remained unchanged during the three days. Therefore, the data in 2017/1/1~2017/1/3 is regarded as abnormal data and all eliminated in the data set.

After removing the three-day unchanged operation data, the data set is normalized, and the box diagram is drawn as shown in fig.5. Boxplot is an algorithm that uses five statistics in data, e.g., minimum, first quartile, median, third quartile and maximum, to analyze data distribution. It can be used to represent the distribution of data and analyze whether it is symmetric and can be used to quickly identify outliers[25].

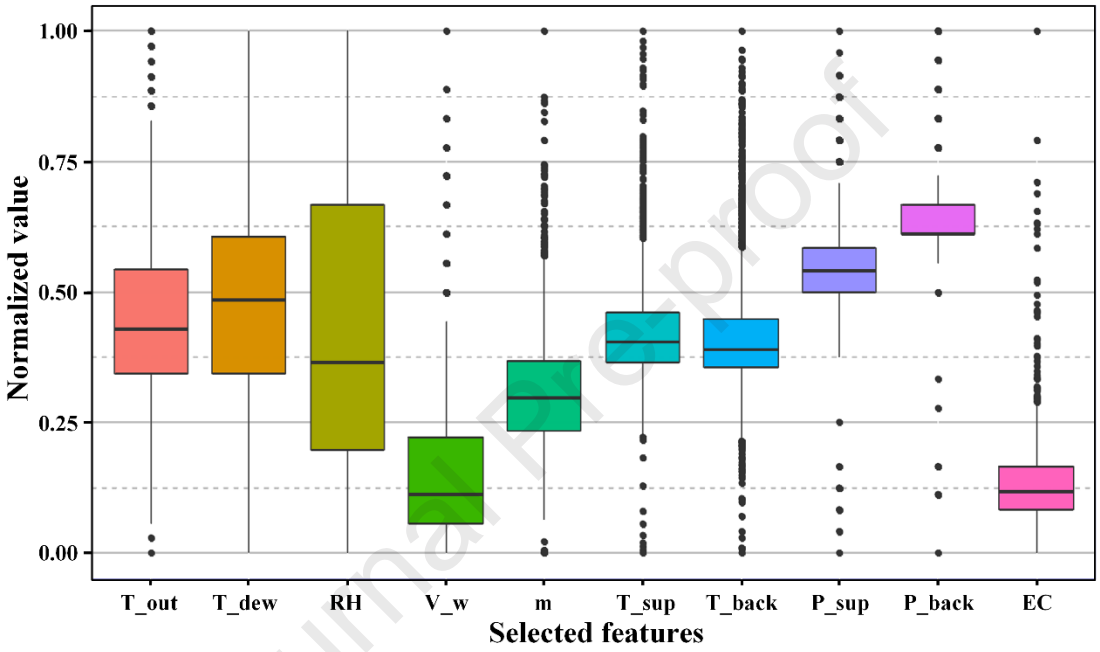



Fig.5 Original data variable distribution

As shown in fig.5, the black dots in the figure indicate abnormal values. There are abnormal values in outdoor temperature, wind speed, instantaneous flow, water supply temperature, backwater temperature, water supply pressure, backwater pressure, and instantaneous heat, which contain 80, 12, 55, 264, 320, 18, 12, and 89 abnormal values, respectively. According to box plot analysis, 426 samples were diagnosed as abnormal, accounting for about 7.5% of the total data.

Because the abnormal value of each variable is detected individually in boxplot analysis, not all the abnormal values need to eliminate. However, containing an

abnormal value makes the sample invalid. Thus samples with any abnormal variable value will be considered abnormal samples and processed, e.g., eliminated or corrected. The samples are eliminated if the samples are continuously abnormal for an extended period. If the abnormal appears discretely or only lasts for a brief period, the abnormal data will be replaced by interpolation to correct the sample based on the previous data before and after.

### 3.2.2 Feature selection

Feature selection is an indispensable step in data analysis. Selecting proper feature variables for modeling can improve the performance of the model. It can reduce the dimensions of the data set by reducing the number of features, thus improving the model's generalization ability and reducing the risk of over-fitting, and making the characteristics of data and the running process of the system more uninterpretable. This section will feature the selection of the preprocessed data based on the maximum information coefficient (MIC) and Bortua algorithm.

MIC is based on the mutual information coefficient; MIC(X, Y) is expressed as the percentage of information that Y can be explained by variables X. The maximum information coefficient of the variables X and Y are MIC(X, Y)=1. Ideally, when two variables are independent, MIC(X, Y)=0. MIC is used to analyze variables in a large amount of closely related data, and it is a criterion to represent the correlation between two variables[26]. It has an advantage in fairness and standardization.

Fig.6 is the matrix of MIC among variables. The lower left part of the matrix is the maximum information coefficient value, the upper right part is the elliptic graph

reflecting the maximum information coefficient, and the maximum information coefficient value on the diagonal line is 1. The smaller the area of the upper right corner ellipse or the darker the color, the larger the maximum information coefficient between the corresponding two variables. The order of input variables and predicted energy consumption coefficients is: instantaneous heat, instantaneous flow rate, water supply temperature, return water temperature, return water pressure, outdoor temperature, water supply pressure, relative humidity, wind speed, and dew point temperature. The dew point temperature, wind speed, and the coefficient of predicted energy consumption are all less than 0.1, which are regarded as invalid variables. The characteristic sets of input variables are instantaneous heat, instantaneous flow rate, water supply temperature, backwater temperature, backwater pressure, outdoor temperature, water supply pressure, and relative humidity.

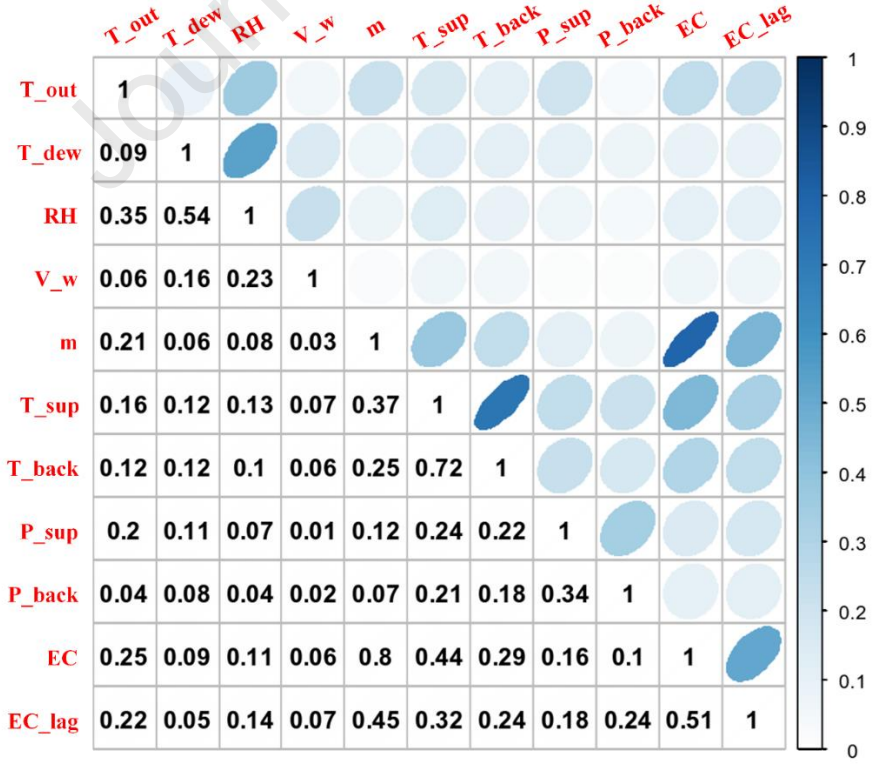



Fig.6 Maximum information coefficient map between variables

Boruta algorithm is a wrapper based on the Random Forest algorithm. Random forest is a classification method based on the voting of multiple unbiased weak classifier decision trees. The weak classifier decision trees are trained independently on different training samples, and the importance of each variable in the final classifier is analyzed comprehensively. The difference between the Bortua algorithm and random tree is that it considers the fluctuation of model accuracy loss in the analysis of weak classifiers and uses average descent accuracy to judge the importance of each variable[27].

When the Boruta algorithm is applied, the order of each real feature variable R is randomly shuffled to obtain the shadow feature matrix S, which is then connected to the actual feature to form a new feature matrix N = [R, S]. The new matrix N is used to train the model to get the feature importance of actual features and shadow features. If the score of the variable in the original data is higher than the score of the shadow feature, the feature gets a score in the importance evaluation. According to the setting of the number of iterations n, the algorithm will randomly arrange the shadow features and evaluate each variable n times to obtain the variables with high importance according to the feature importance score in the evaluation result.

The importance of variables obtained by the Boruta algorithm is shown in fig.7. Blue boxed graphs represent the minimum, average and maximum z-scores of a shadow attribute, while green boxed graphs represent confirmed attributes. The result shows that the eight feature variables selected are considered important variables by comparison. The order of importance from high to low is as follows: $T_{out}$, $T_{sup}$, $T_{back}$, EC, $P_{sup}$, m, RH and $P_{back}$.

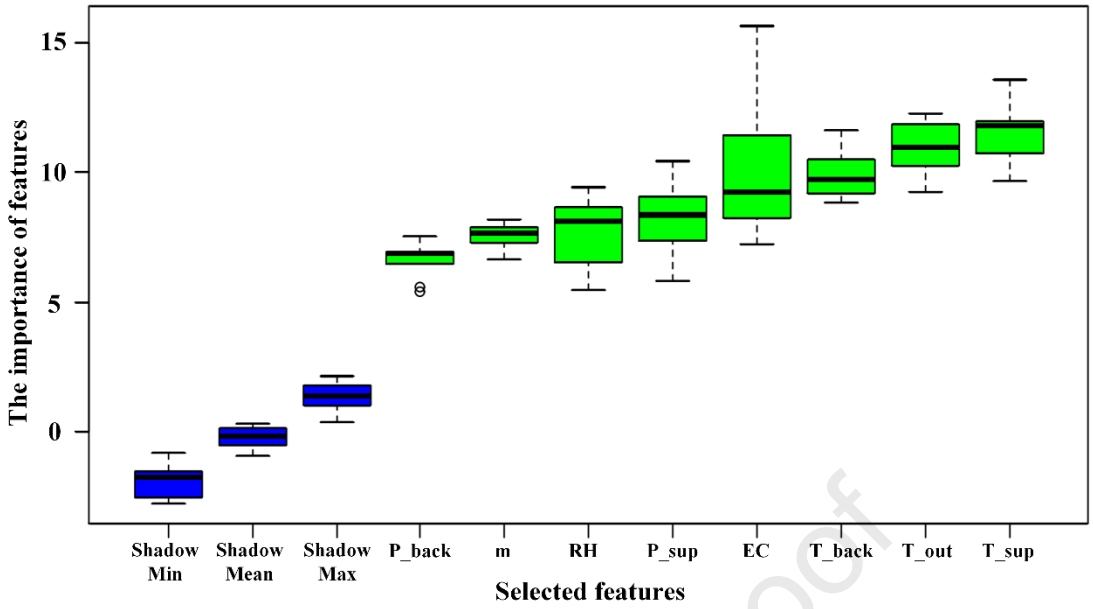



Fig.7 Sorting Graph of Importance of Variables

This research uses recursive feature elimination to select the best number of feature variables. The main idea of recursive feature elimination is to establish models repeatedly with different input feature variables. The decision tree model is used in this paper, and the best feature variables are selected according to the RMSE in the evaluation result of the models. The feature variables with the highest importance are selected from the initial data set of feature variables and added to the model's input; the process is repeated on the remaining features until all the features are traversed[28]. Decision tree models with different input feature variables are constructed, and the RMSE of each model is evaluated and compared.

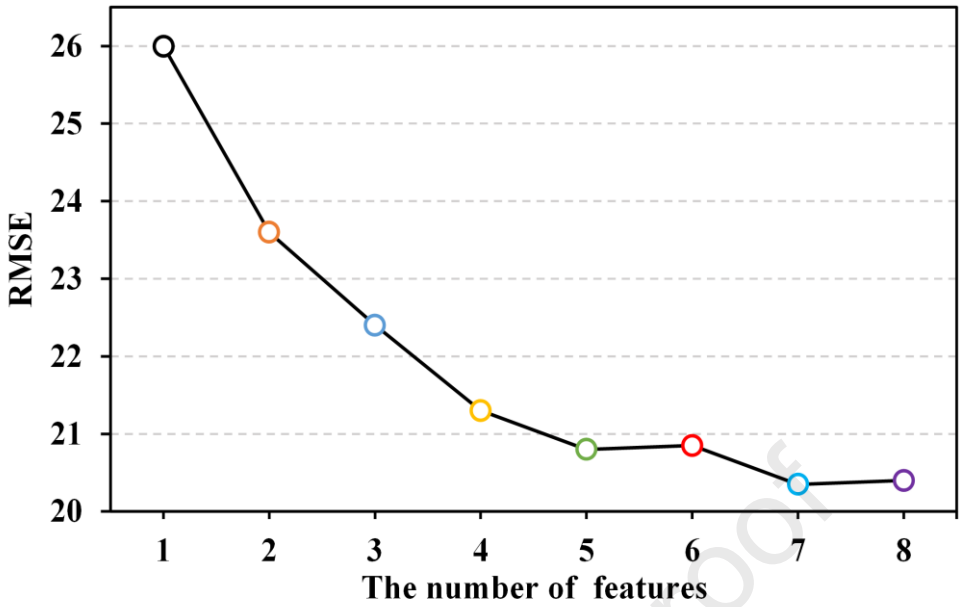



Fig.8 Number of features and RMSE diagram

As shown in fig.8, the number best number of input feature variables is seven, as the prediction model has the highest accuracy. Thus the final input set of the model is [$T_{out}$, $T_{sup}$, $T_{back}$, EC, $P_{sup}$, m, RH].

### 3.2.3 Creating a new feature

In resident buildings, the heating energy consumption can be significantly influenced by historical energy consumption [29], and the energy consumption usually has good continuity. Existing study also mentioned that past time-steps energy consumption could account for the potential impact of past events on the current and predicted states of the building energy [30]. Therefore, a new feature is extracted by EWMA variable to improve the accuracy of energy consumption prediction. The new feature is created based on the following equation:

$$EWMA_i = \begin{cases} \mu_i \cdot (1-\lambda) + x_i \cdot \lambda & i = 1 \\ EWMA_{i-1} \cdot (1-\lambda) + x_i \cdot \lambda & i \geq 2 \end{cases} \quad (4)$$

In this research, $\mu_i(i=1)$ is the energy consumption at the beginning, 8:30 am, on

November 15, 2016. $x_i$ is the heating energy consumption at the i-th moment. The historical comprehensive energy consumption index at the i-th moment is marked as $EWMA_i$, while λ denotes an attenuation coefficient representing previous energy consumption's influence.

## 4. Results and discussion

After the data preprocess and feature selection, the energy consumption of actual data of the research object in this paper is shown in fig.9. The data set is divided into a training set and a test set in a ratio of 7:3 for all the proposed models. In fig.9, the curve left to the dotted line is the energy consumption curve of the train set, and the right side is the energy consumption curve of the test set. The input feature variables and the corresponding values are shown in table 2. The fitting formula of the model is as follows:

Cumulative heat of predicted time interval =f($T_{out}$, RH, $T_{sup}$, $T_{back}$, $P_{sup}$, $m$, $EC$, EWMA)

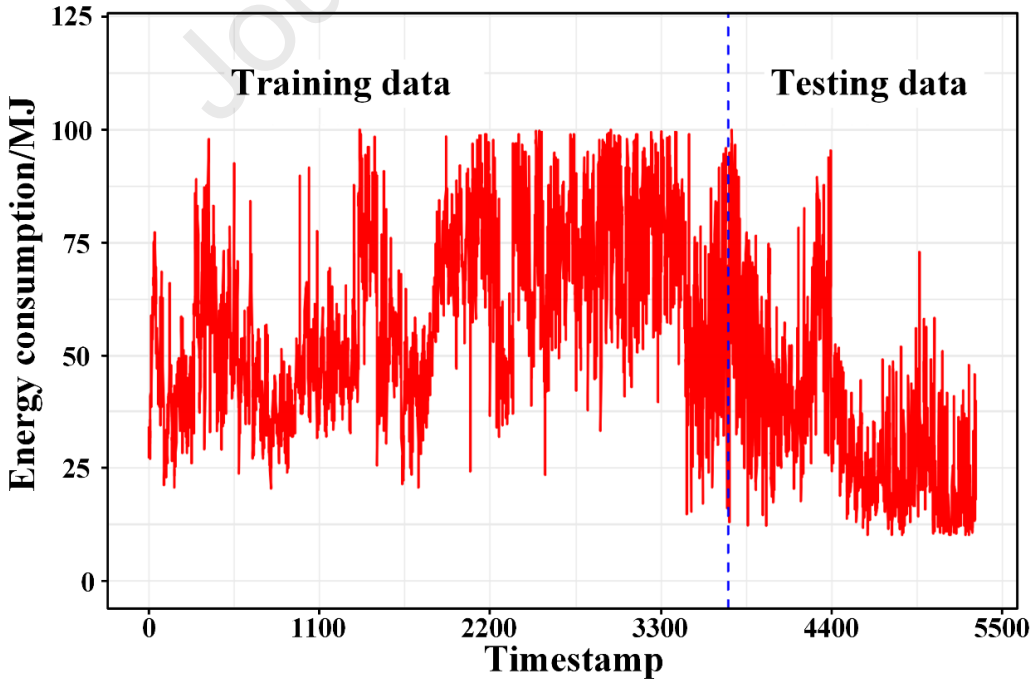



Fig.9 Curve chart of actual energy consumption value change

Table.2 Detailed Modeling Variables

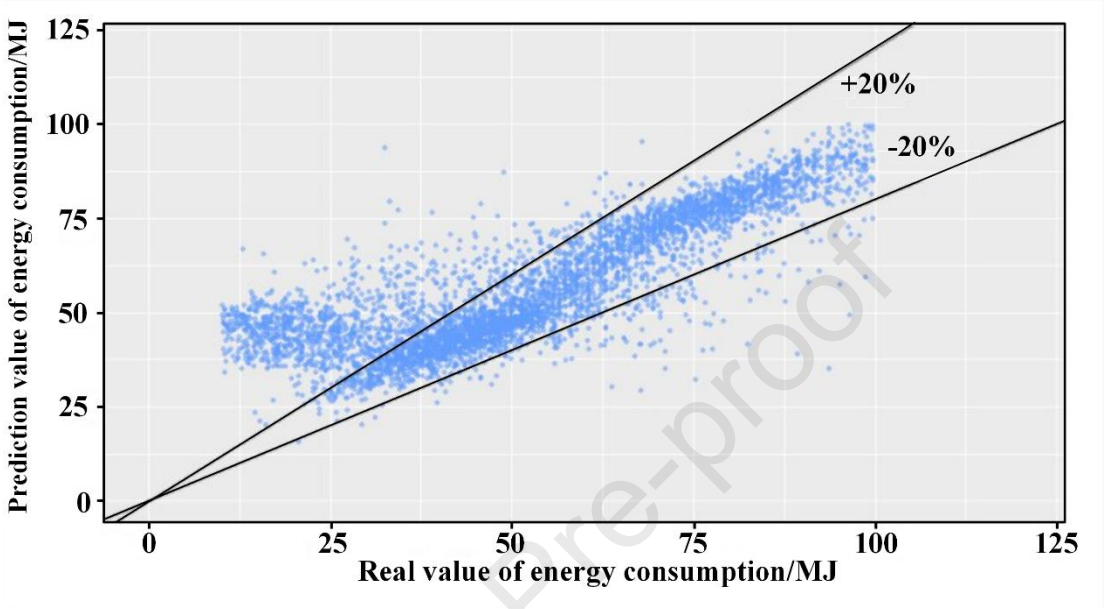
| Variable category | Variable name | Variable range |
|---|---|---|
| **Meteorological parameters** | $T_{out}$ (℃) | -15 ~ 20 |
| | RH | 0.04 ~ 1.00 |
| **Operation parameters** | $T_{sup}$ (℃) | 37.30 ~ 63.90 |
| | $T_{back}$ (℃) | 32.10 ~ 57.70 |
| | $P_{sup}$ $(kPa)$ | 0.84 ~ 1.07 |
| | $m$ $(m^3/h)$ | 1.50 ~ 24.00 |
| | $EC$ $(MJ)$ | 4.54 ~ 242.55 |
| | $EWMA$ $(MJ)$ | 7.56 ~ 225.34 |
| **Prediction variable** | $EC\_lag$ $(MJ)$ | 4.54 ~ 244.82 |

This research uses MLR, ELM, XGB, and SVG algorithm to establish prediction models for resident building energy consumption, and the four base machine learning algorithms are combined to form the ensemble model. The result of each performance evaluation will be shown in this section, and the influence of using EWMA feature on the ensemble model will be analyzed and discussed.

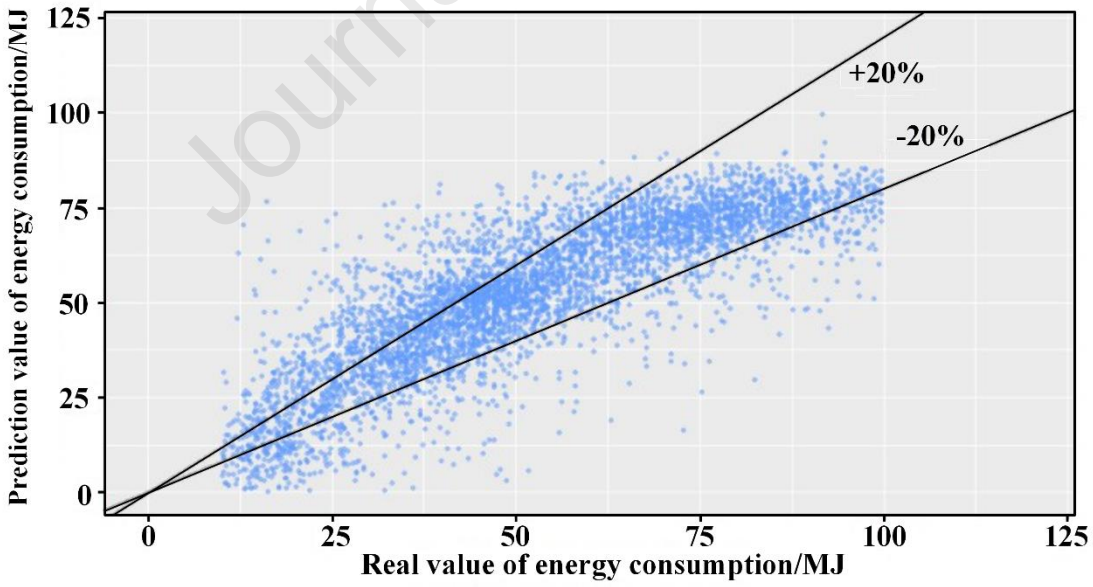
## 4.1 Base models

According to the four base models trained on the test set, energy consumption on the test set is predicted and compared with the actual value of energy consumption, where all input parameters were used to train each base models. Fig.10 compares the predicted and actual values of building energy consumption in four months. The abscissa represents the actual value of building energy consumption, and the ordinate

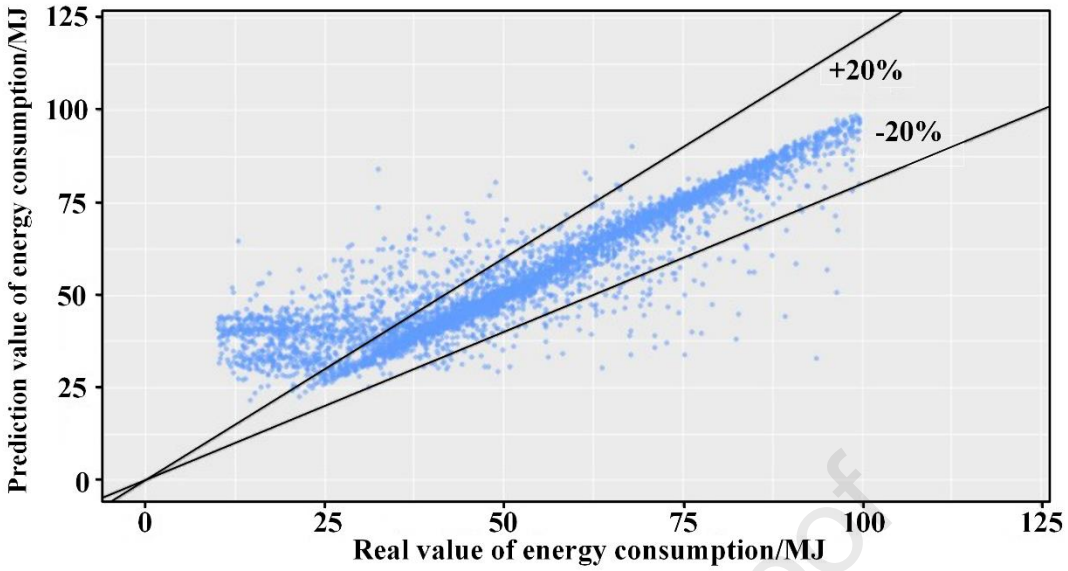
represents the predicted energy consumption value. Each base model is trained and evaluated ten times independently, and the data in the graph is according to the average value of prediction results.



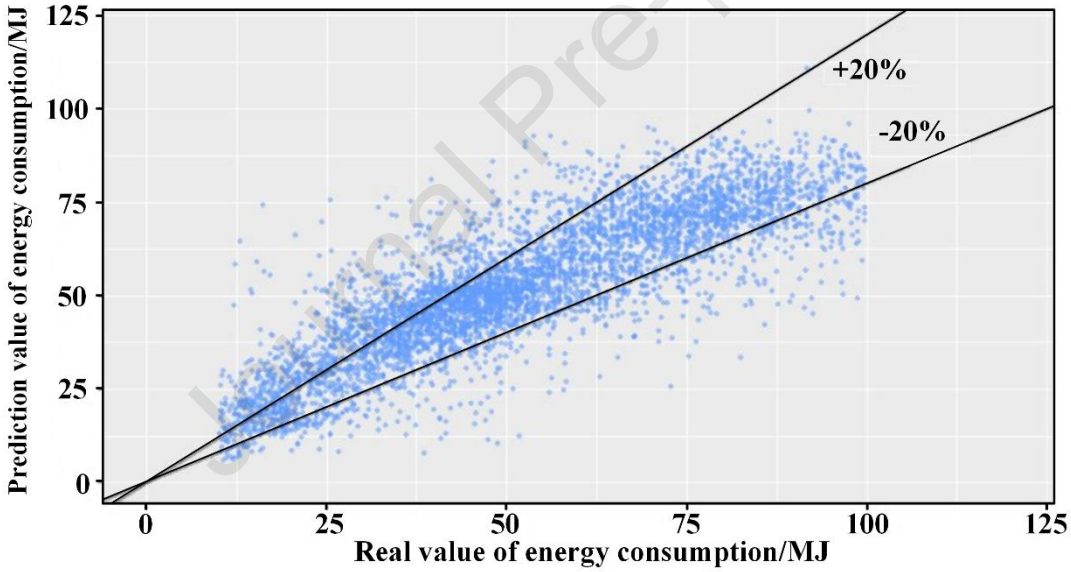

(a) MLR




(b) ELM




(c) XGB




(d) SVR

Fig.10 Predicted energy consumption map of base models

In the comparison result, the proportion of samples whose real value is within the ±20% deviation range of the predicted value is calculated for each base model. The proportion samples within ±20% deviation of MLR, ELM, XGB and SVR is 69.6%, 76.9%, 78.9% and 76.2%, respectively. The MAE and RMSE of predicted and actual

values are summarized in Table 3. According to the evaluation result, the order of base models with the MAE on testing data from small to large is XGB, ELM, SVR and MLR, and that with RMSE is XGB, SVR, ELM and MLR.

Table.3 Detailed Prediction results of base models

| Model | MLR | | ELM | | XGB | | SVR | |
|---|---|---|---|---|---|---|---|---|
| Data set | Training set | Testing set | Training set | Testing set | Training set | Testing set | Training set | Testing set |
| MAE | 8.48 | 14.56 | 8.56 | 9.53 | 7.97 | 8.85 | 8.96 | 9.62 |
| RMSE | 11.28 | 16.95 | 11.40 | 12.02 | 10.71 | 11.75 | 11.16 | 11.89 |

As a matter of fact, the training data contains more high energy consumption data but less low energy consumption data during winter. Thus, it's obvious that all four base models show more deviation on predicting low energy consumption, which means base models have poor generalization ability. The XGB model, as a boosting ensemble decision tree model, has the lowest MSE and RMSE not only on training data but also testing data. Although the prediction distribution showed that XGB could have higher accuracy on high energy consumption prediction, especially over 50MJ, yet the XGB shows worse deviation on low energy consumption prediction, and this result is similar to that of MLR model. Meanwhile, the biggest accuracy difference of MLR between training and testing data could also prove that the XGB and MLR models are easier to overfit than ELM and SVR models.

It's noticeable that the energy consumption prediction distribution of ELM and SVR models is more uniform than that of MLR and XGB models, and the similar MSE

and RMSE indexes between ELM and SVR model also indicated that these two models could have similar energy prediction performance, even the learning structure of them is totally different with each other. It's noteworthy that the ELM and SVR models will not predict low energy consumption as high energy consumption dramatically. Thus, the SVR and ELM models could have better generalization ability than MLR and XGB models. This phenomenon also indicates that ensemble model utilizing above-mentioned four base models has the potential to balance the ability to high prediction accuracy and better generalization.

## 4.2 Ensemble model

The ensemble prediction model is established according to the process in section 3 and is evaluated. The predicted energy consumption of the ensemble model is compared with the actual value of the ensemble learning model, and Fig.11 shows the comparison result. Meanwhile, to highlight the performance of ensemble model, the ensemble model was compared with the best base XGB model and all input parameters were used to train ensemble model as well. The RMSE and MAE in the evaluation result of the ensemble learning algorithm model and XGB model are shown in table 4. In the comparison result, 82% of all data are located in the ±20% deviation range of actual energy consumption, which is 3.9% higher than the XGB model. Compared with the prediction results of XGB, although the ensemble model didn't perform better than XGB model on training data set, yet the MAE of the ensemble model is reduced by 4.36% on the test set, and the RMSE is reduced by 3.80% on the test set. These results indicated that ensemble model could reduce overfitting efficiently. In addition, the

distribution of predicted data from ensemble model was more uniform compared with this from XGB, which also proved that ensemble model could pay more attention to overall energy consumption features but not local features.

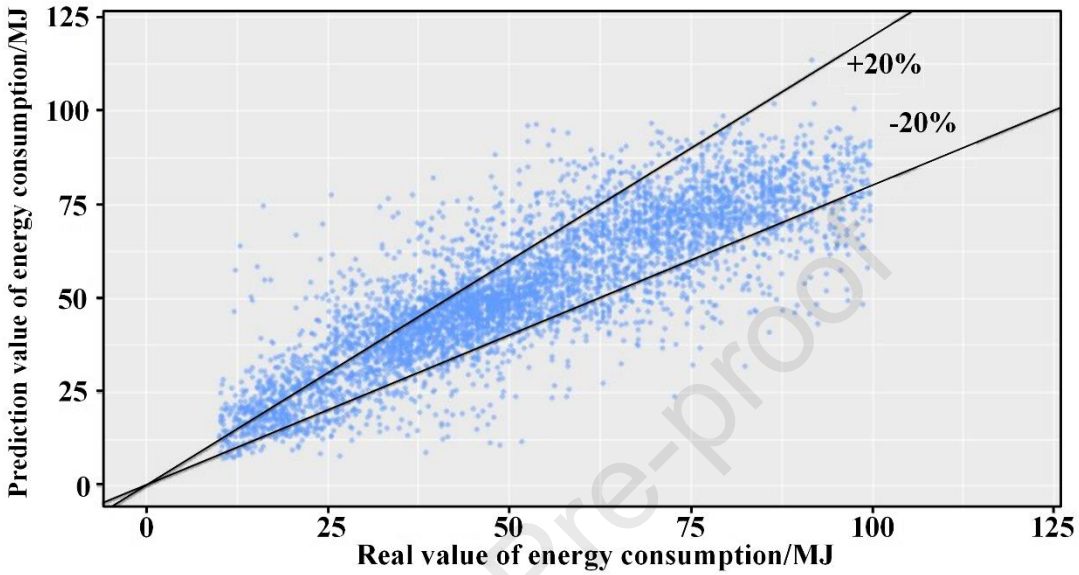



Fig.11 Predicted energy consumption map of ensemble learning

Table.4 Comparison table between ensemble learning and XGB

| Model | Ensemble learning | | XGB | |
|---|---|---|---|---|
| Data set | Training set | Testing set | Training set | Testing set |
| MAE | 8.39 | 8.48 | 7.97 | 8.85 |
| RMSE | 10.86 | 11.32 | 10.71 | 11.75 |

## 4.3 Influence of EWMA feature

In order to analyze the influence of the historical energy consumption feature variable EWMA, an ensemble learning model without EWMA is established and evaluated. Fig.12 shows the comparison of the predicted value of energy consumption and the actual value of the ensemble model without EWMA. 75.4% of all data are located in the ±20% deviation range of actual energy consumption, which is 8.1% less

than the ensemble model with EWMA. Another obvious difference is that the ensemble model without EWMA feature is easier to regard low energy consumption as high energy consumption, and the prediction distribution is not as uniform as that of ensemble model with EWMA feature. That also indicates that historical energy consumption feature could improve the prediction generalization ability to some extent. As shown in Table 5, the MAE and RMSE of ensemble model with EWMA feature are reduced by 10.36% and 19.89% respectively, on the test set compared with to ensemble model without EWMA, where the improvement is better than that between the ensemble model and XGB model. The ensemble model with EWMA feature not only shows higher accuracy on both training and testing data but also more uniform prediction deviation.

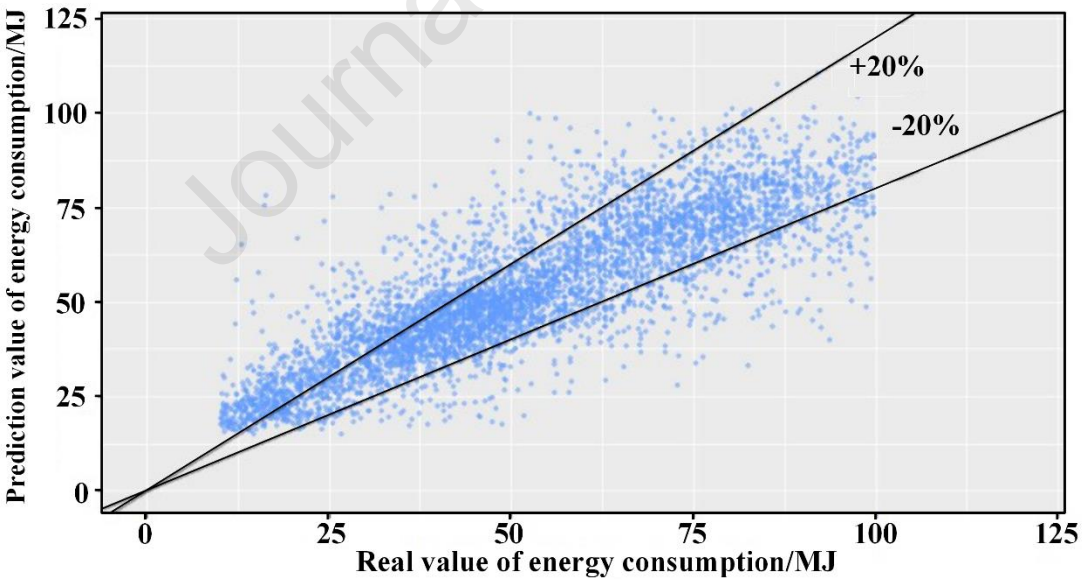



Fig.12 Predicted energy consumption map of ensemble learning without the EWMA

Table.5 Comparison between ensemble learning with and without EWMA

| Model | Ensemble learning | Ensemble learning without EWMA |
|---|---|---|

| Data set | Training set | Testing set | Training set | Testing set |
|---|---|---|---|---|
| MAE | 8.39 | 8.48 | 9.83 | 9.46 |
| RMSE | 10.86 | 11.32 | 11.62 | 14.13 |

## 5. Conclusions

This paper is based on operation data of a heating system in a residential district. base machine learning and ensemble learning algorithms are used to establish an energy consumption prediction model based on unit operation data, local meteorological data, and the total energy consumption of the resident district. The ensemble learning model is optimized by abnormal value processing, feature selection, and parameter optimization, and finally, the ensemble learning model is established. The ensemble model is compared with several base models and a critical feature is evaluated. The conclusions of this paper are as follows:

(1) The abnormal data is eliminated based on boxplot analysis, and it is found that backwater temperature has the most detected abnormal values. There are 426 samples eliminated and interpolation is used to make up eliminated data. In feature selection process, the maximum information coefficient and Bortua algorithm are combined to analyze the contributions of different input features, where instantaneous heat and wind speed show the most and least connection with energy consumption respectively. The recursive feature elimination is used to select seven critical features for energy consumption prediction.

(2) A stacking ensemble model based on four base machine learning methods, e.g., MLR, ELM, XGB, and SVR, is built to predict resident building energy consumption.

It's noticeable that XGB model, as a boosting ensemble decision tree model, has the best prediction performance among four base models. Compared with the best base XGB model, the ensemble model shows lower accuracy on training data set, but it performs better on testing data set, where the MAE of the ensemble model is reduced by 4.36% and the RMSE is reduced by 3.80% on the test data set. The results indicate that the stacking ensemble model could reduce overfitting of energy consumption prediction to some extent.

(3) The impact of historical energy consumption on prediction performance is evaluated. A new feature based on exponentially weighted moving average is added to train ensemble model. Compared with ensemble model without historical feature, the proportion of samples within the ±20% deviation range of actual energy consumption increases by 8.8%. In addition, the historical feature reduces the MAE on the test set by 10.36%, and the RMSE by 19.89%. The comparison between ensemble model with the historical feature and without the historical feature indicates that historical energy consumption data could improve prediction accuracy dramatically.

Although the stacking ensemble model shows the best performance on predicting resident building energy consumption compared with four common methods, yet the generalization of proposed method should be evaluated by testing method with different building energy consumption prediction tasks, which should be studied in future work.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China. (No. 51876070)

# References:


[1]. Bianchini, G., et al., An integrated model predictive control approach for optimal HVAC and energy storage operation in large-scale buildings. Applied Energy, 2019. 240: p. 327-340.

[2]. Shabunko, V., C.M. Lim and S. Mathew, EnergyPlus models for the benchmarking of residential buildings in Brunei Darussalam. Energy and Buildings, 2018. 169: p. 507-516.

[3]. Bansal, N.K. and M.S. Bhandari, Comparison of the periodic solution method with TRNSYS and SUNCODE for thermal building simulation. Solar Energy, 1996. 57(1): p. 9-18.

[4]. Luo, C., C. Tan and Y. Zheng, Long-term prediction of time series based on stepwise linear division algorithm and time-variant zonary fuzzy information granules. International Journal of Approximate Reasoning, 2019. 108: p. 38-61.

[5]. Gorgannejad, S., et al., Quantitative prediction of the aged state of Ni-base superalloys using PCA and tensor regression. Acta Materialia, 2019. 165: p. 259-269.

[6]. Li, K., et al., Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. Energy and Buildings, 2015. 108: p. 106-113.

[7]. Gong, M., et al., Heat load prediction of residential buildings based on discrete wavelet transform and tree-based ensemble learning. Journal of Building Engineering, 2020. 32: p. 101455.

[8]. Ahmad, T., H. Zhang and B. Yan, A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society, 2020. 55: p. 102052.

[9]. Wang, Z. and R.S. Srinivasan, A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renewable and Sustainable Energy Reviews, 2017. 75: p. 796-808.

[10]. Li, Q., et al., Applying support vector machine to predict hourly cooling load in the building. Applied Energy, 2009. 86(10): p. 2249-2256.

[11]. Golkarnarenji, G., et al., Support vector regression modelling and optimization of energy consumption in carbon fiber production line. Computers & Chemical Engineering, 2018. 109: p. 276-288.

[12]. Wang, Z., Y. Wang and R.S. Srinivasan, A novel ensemble learning approach to support building energy use prediction. Energy and Buildings, 2018. 159: p. 109-122.

[13]. Ofori-Ntow Jnr, E., Y.Y. Ziggah and S. Relvas, Hybrid ensemble intelligent model based on wavelet transform, swarm intelligence and artificial neural network for electricity demand forecasting. Sustainable Cities and Society, 2021. 66: p. 102679.

[14]. Xuan, W., et al., A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. International Journal of Electrical Power & Energy Systems, 2021. 126: p. 106583.

[15]. Chen, K., et al., A novel data-driven approach for residential electricity consumption prediction based on ensemble learning. Energy, 2018. 150: p. 49-60.

[16]. Polikar, R., Ensemble Learning, in Ensemble Machine Learning: Methods and Applications, C. Zhang and Y. Ma, C. Zhang and Y. Ma^Editors. 2012, Springer US: Boston, MA. p. 1-34.

[17]. Jiménez, M.J. and M.R. Heras, Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing. Solar Energy, 2005. 79(3): p. 302-310.

[18]. Huang, G., Q. Zhu and C. Siew, Extreme learning machine: Theory and applications. Neurocomputing, 2006. 70(1): p. 489-501.

[19]. Fan, J., et al., Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conversion and Management, 2018. 164: p. 102-111.
[20]. Xu, L., et al., Representative splitting cross validation. Chemometrics and Intelligent Laboratory Systems, 2018. 183: p. 29-35.
[21]. Bhat, P.C., et al., Optimizing event selection with the random grid search. Computer Physics Communications, 2018. 228: p. 245-257.
[22].Qi, J., et al., On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. IEEE SIGNAL PROCESSING LETTERS, 2020. 27: p. 1485-1489.
[23]. Ćalasan, M., S.H.E. Abdel Aleem and A.F. Zobaa, On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function. Energy Conversion and Management, 2020. 210: p. 112716.
[24]. Lu, Y., et al., GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system. Energy and Buildings, 2019. 190: p. 49-60.
[25]. Lem, S., et al., The heuristic interpretation of box plots. Learning and Instruction, 2013. 26: p. 22-35.
[26]. Sun, G., et al., Feature selection for IoT based on maximal information coefficient. Future Generation Computer Systems, 2018. 89: p. 606-616.
[27]. Amiri, M., et al., Assessment of the importance of gully erosion effective factors using Boruta algorithm and its spatial modeling and mapping using three machine learning algorithms. GEODERMA, 2019. 340: p. 55-69.
[28]. Sahran, S., et al., Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading. Artificial Intelligence in Medicine, 2018. 87: p. 78-90.
[29]. Somu, N., G. Raman M R and K. Ramamritham, A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 2021. 137: p. 110591.
[30]. Bourdeau, M., et al., Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities and Society, 2019. 48: p. 101533.

## Highlights:

- A case study on a heating station system in Beijing was illustrated.
- A novel heat supply energy consumption prediction strategy was proposed.
- A stacking ensemble model was trained and verified in this study.

AUTHOR STATEMENT:

Zhang Jianxin: Conceptualization, Investigation, Visualization, Writing-Review & Editing.

Huang Yao: Conceptualization, Methodology, Software, Writing-Original Draft.

Cheng Hengda: Investigation, Data curation, Resources.

Chen Huanxin: Funding acquisition, Supervision.

Xing Lu: Writing-Review & Editing.

He Yuxuan: Validation, Writing-Review & Editing

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: