International Journal of Economy, Energy and Environment
2025, Vol. 10, No. 3, pp. 65-71
https://doi.org/10.11648/j.ijeee.20251003.12

Research Article

A Stacking-Based Ensemble Model for Short-Term Load Forecasting

Kecheng Huang[1], Yidong Yu[2, *], Kang Jiang[2]

[1]Emerging AI Technology Co. Ltd., Hangzhou, China

[2]School of Information Technology and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China

Abstract

Short-term load forecasting plays an important and indispensable role in the daily operation planning of power grid because it allows grid operators to predict electricity demand a few hours to one week in advance. Although statistics-based methods and machine learning-based methods have been widely used in short-term load forecasting, a single model may have difficulty capturing all underlying dynamics, causing reduced prediction accuracy. Therefore, a stacking-based ensemble model that improves prediction accuracy by integrating multiple base prediction models is proposed in this study for short-term load forecasting. Firstly, for data preprocessing, data normalization is used to scale the raw load data to a range of 0 to 1. Data imputation is used to ensure data integrity. Secondly, base prediction models including logistic regression, decision tree, random forest, multilayer perceptron, convolutional neural network, and long short-term memory are utilized to train the prediction models. Thirdly, the stacking-based ensemble learning method is utilized to integrate these base prediction models to further predict electric load. The results of comparative experiments and error analysis show that the stacking-based ensemble learning model outperforms the base prediction models for the majority of the evaluation metrics. Additionally, the analysis of curve fitting results demonstrates the high level of agreement between the actual values and the predicted values for the stacking-based ensemble learning model.

Keywords

Load Forecasting, Stacking, Ensemble Learning Method

1. Introduction

Short-term load forecasting (STLF) is of great importance for the daily operation planning of power gird, as it enables grid operators to accurately predict electricity demand over periods ranging from a few hours to one week. By providing reliable estimates of future load variations, STLF supports critical decision-making processes such as economic dispatch, unit commitment, contingency analysis, and reserve allocation. STLF ensures the balance between electricity demand and supply, minimizes operational costs, improves the reliability of the power grid, and facilitates the integration of renewable energy sources by anticipating their intermittent nature [1].

The majority of existing STLF models leverage two primary methods: statistics-based methods and machine learning-based methods [2]. Statistics-based methods, such as

*Corresponding author: 1121395432@zufe.edu.cn (Yidong Yu)

Received: 14 April 2025; **Accepted:** 15 May 2025; **Published:** 11 June 2025

autoregressive integrated moving average (ARIMA) [3], exponential smoothing (ES) [4], and multiple linear regression (MLR) [5], rely on historical load data and mathematical formulations to capture temporal patterns, seasonality, and trends. These models are widely adopted due to their interpretability, computational efficiency, and well-established theoretical foundations. However, they often struggle with nonlinear relationships and complex load patterns influenced by external factors like weather and holidays [6].

On the other hand, machine learning-based methods, including artificial neural network (ANN) [7], support vector machine (SVM) [8], random forest (RF) [9], and more recently deep learning models, excel at handling nonlinear relationships and learning complex load patterns from large datasets, which alleviates the shortcomings of the statistics-based methods [10]. However, STLF is influenced by multiple complex and interdependent factors which introduce high nonlinearity and uncertainty into load patterns. Therefore, a single model may struggle to fully capture all underlying dynamics, leading to reduced prediction accuracy.

To overcome the limitations mentioned above, a stacking-based ensemble learning model (SBEM) which enhances prediction accuracy by combining multiple base prediction models is proposed in this study for STLF. Firstly, electric load data is preprocessed via normalization and imputation. Secondly, the base prediction models including logistic regression (LR) [11], decision tree (DT) [12], random forest (RF) [9], multilayer perceptron (MLP) [13], convolutional neural network (CNN) [14], and long short-term memory (LSTM) [15] are used to train an ensemble learning model and predict future electric load. Thirdly, these single base prediction models are integrated to predict electric load with a stacking-based ensemble learning method. The comparative experiments and error analysis between these single base prediction models and the proposed model validate the effectiveness of the SBEM.

The remainder of the study is organized as follows. Section 2 reviews existing research on statistics-based methods and machine learning-based methods. Section 3 explores the proposed SBEM. Experimental comparisons and results analysis are shown in Section 4. Finally, Section 5 concludes the study and discusses future research.

2. Related Work

This section conducts a review of the existing research works regarding statistics-based methods and machine learning-based methods to offer a more in-depth comprehension of the previous research.

2.1. Statistics-Based Methods

ARIMA is a widely used statistics-based method, which can capture trends, seasonality, and random fluctuations in the load data. Lee and Ko [3] proposed a model combining lifting scheme and ARIMA for STLF. Specifically, the lifting scheme breaks down the initial load series into sub-series, which are then forecasted using ARIMA models. In order to fairly compare direct and indirect ARIMA-based approaches, Shi et al. [16] conducted a comparative experiment on wind speed and power generation data from an offshore wind turbine. The results show that the direct approach outperforms the indirect approach in terms of root mean square error and mean absolute error. Although ARIMA can effectively analyze stationary time series, it has strict requirements for data stationarity and a complex parameter determination process.

In contrast, ES has lower requirements for stationarity, which is based on weighted averages with exponentially decaying weights. Göb et al. [4] explored the use of meteorological covariates in STLF via the ES with covariates. The empirical study on Italian electricity consumption data shows that the proposed model has an excellent forecasting precision, especially for prediction horizons of 1 day and more. MLR is a statistics-based method that establishes a model to depict the relationship between a dependent variable and several independent variables. Compared with ARIMA and ES, MLR is able to incorporate various influencing factors easily. Fang and Lahdelma [5] evaluated MLR and seasonal ARIMA for heat demand forecasting in a district heating system. MLR considers weather and social factors with parameters optimized by least squares. The results show that MLR has the highest accuracy, outperforming the seasonal ARIMA-combined model and other regression models.

Although statistics-based methods have wide applications in STLF due to their interpretability, computational efficiency, and well-established theoretical foundations, they still struggle with nonlinear relationships and complex load patterns.

2.2. Machine Learning-Based Methods

Machine learning-based methods have emerged as powerful and indispensable tools in STLF, which are good at handling nonlinear relationships and learning complex load patterns from large datasets, addressing the deficiencies of statistics-based methods.

ANN is a powerful and commonly used machine learning-based method, which has a strong nonlinear mapping ability. Hsu and Chen [7] utilized ANN for regional load forecasting in Taiwan. They divided Taiwan into four regions and constructed ANN models with regional gross domestic product (GDP), population, and highest temperature as inputs. The results demonstrate that ANN is more accurate than the traditional regression-based model. SVM is also a popular and effective machine learning-based method, which seeks an optimal hyperplane in the feature space. Dai et al. [8] used SVM with ship speed, wind speed, and water flow velocity as features. The improved particle swarm optimization algorithm was applied to optimize the parameters of SVM. The case simulation shows that the proposed model achieves a good prediction accuracy with a reduced average relative error.

With the continuous expansion of the scale and increase in the complexity of the power system, traditional machine learning-based methods gradually show limitations when dealing with nonlinear, uncertain, and complex dynamic load data. Deep learning-based methods, a subfield of machine learning-based methods, have shown their powerful capabilities in STLF [17]. Kong et al. [15] proposed an LSTM-based framework for STLF. Compared with multiple benchmarks like empirical methods and other machine learning approaches, the LSTM-based framework generally achieves the best performance in individual and aggregated load forecasting. In addition, Kim and Cho [14] proposed a CNN-LSTM neural network for predicting residential electric load consumption. It uses CNN to capture spatial features from multiple variables, and LSTM to model temporal information. Experiments show that it outperforms conventional methods, achieving the lowest error metrics.

Although the machine learning-based methods have been widely used in STLF, they still face the problem that a single model may struggle to fully capture all underlying dynamics, leading to reduced prediction accuracy.

3. Methodology

As shown in Figure 1, in order to address the deficiencies of a single model, a stacking-based ensemble learning model is proposed in this study.

3.1. Data Preprocessing

Load data from different time periods or different systems may vary greatly in terms of value range and time scale, which will bring inconvenience when performing mathematical operations or comprehensive analysis. Therefore, in this study, min-max normalization is employed, which scales the raw load data to the range from 0 to 1. The process of min-max normalization can be described as Eq. (1) [18].

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where x is the raw load data; x_{min} is the minimum of x; x_{max} is the maximum of x; x_{norm} is the normalized load data.

Besides, in the actual operation process of the power system, due to various reasons such as equipment failure, communication interruption, and measurement errors, load data may be missing, which will affect the performance of the proposed model. Therefore, zero is used for imputation to ensure the integrity of data in this study.

3.2. Training of Base Prediction Models

In this stage, six base prediction models with superior predictive capabilities are chosen. These models encompass three traditional machine learning models, namely LR, DT, and RF, along with three deep learning models, specifically MLP, CNN, and LSTM. The training set undergoes random sampling six times, generating six independent training subsets. Each of these subsets is utilized to train a base prediction model separately. Subsequently, the validation set is employed to validate the training outcomes, enabling the acquisition of the optimal parameters for the trained base prediction model. Finally, the trained and optimized base prediction models are obtained.

3.3. Training of Stacking-Based Ensemble Learning Model

In this stage, the optimized base prediction models that have been trained in the preceding stage utilize the validation dataset to derive the predictor variables of the validation dataset. These variables are then combined to form a matrix. Subsequently, the ensemble learning model employs these predictor variables as training inputs, resulting in the development of a trained ensemble learning model. Lastly, the test dataset is fed into the trained optimized base prediction models to obtain the predictor variables of the test dataset. These variables are further processed by the trained ensemble learning model to generate the final prediction outcomes.

4. Experiment

Comparative experiments and error analysis are carried out between the single base prediction models and the proposed SBEM in this section. Additionally, the performance of the SBEM is thoroughly analyzed.

4.1. Experimental Settings

All the experiments were executed using Python on a personal computer. The specifications of this computer include an Intel Core i5 CPU running at 1.60GHz, 16GB of Random Access Memory (RAM), and an Intel(R) UHD Graphics GPU.

4.2. Evaluation Metrics

To evaluate the performance of the prediction models, five evaluation metrics are chosen as follows: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), the coefficient of determination (R2), and running time (RT).

MAE: MAE is employed to evaluate the difference between the true value and the predicted value, which can be calculated according to Eq. (2).

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|y_i - \hat{y}_i\right| \quad (2)$$

where N indicates the total number of samples; n indicates the timestep; y_i indicates the true value; $\hat{y}$ indicates the predicted value.

RMSE: RMSE is utilized to measure the extent of deviation between the actual value and the predicted value. RMSE exhibits a higher sensitivity to the outliers present within the dataset and can be computed using Eq. (3).

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(y_i - \hat{y}_i\right)^2} \quad (3)$$

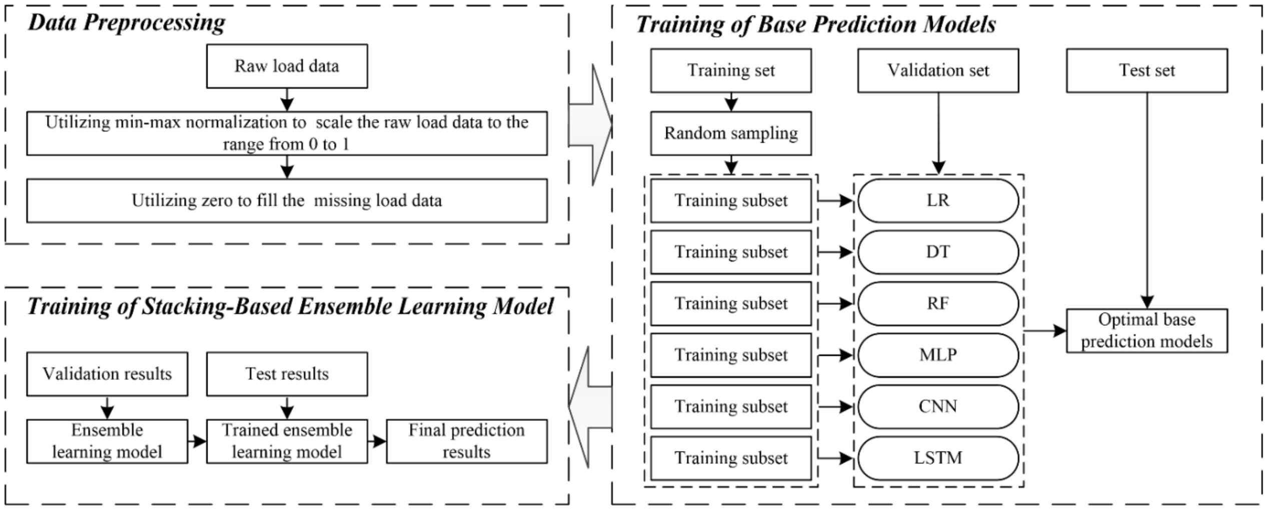


Figure 1. *The structure of the SBEM.*

MAPE: MAPE indicates the average proportion of the relative error existing between the actual value and the predicted value, and MAPE can be determined by using Eq. (4).

$$MAPE = \frac{100\%}{N}\sum_{n=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (4)$$

R2: R2 is employed to assess how well the predictive model fits the data. Contrary to what one might expect, the closer the value of R2 is to 1, the better the model fits the data, signifying a stronger predictive ability. R2 can be computed using Eq. (5).

$$R^2 = 1 - \frac{\sum_{n=1}^{N}\left(y_i - \hat{y}_y\right)^2}{\sum_{n=1}^{N}\left(y_i - \bar{y}_i\right)^2} \quad (5)$$

where $\bar{y}_i$ is the average of the true values.

RT: RT indicates the run time of a model.

The closer the values of MAE, RMSE, MAPE and RT are to zero, the better the model is. Conversely, the closer the value of R2 is to 1, the more superior the model is.

4.3. Comparative Experiments

To evaluate how effectively the proposed SBEM can predict electric load, comparative experiments are carried out. These experiments compare six individual base prediction models with the ensemble learning models. Then, the five previously mentioned evaluation metrics are utilized to characterize the prediction capabilities of these models. The outcomes of comparative experiments are shown in Table 1.

As depicted in Table 1, for the majority of the evaluation metrics, the prediction outcomes of the SBEM surpass those of the individual base prediction models. The results show that when it comes to prediction accuracy and stability, the SBEM demonstrates superiority over the single base prediction models. The SBEM exhibits more excellent performance in the prediction of electric load.

Table 1. *Results of comparative experiments for all models.*

Models	MAE ↓	RMSE ↓	MAPE (%) ↓	R^2 ↑	RT ↓
DT	4549	6604	0.055	0.94	10.13
RF	5331	7005	0.066	0.94	20.90
LR	3779	5626	0.046	0.92	68.14
CNN	3460	5273	0.042	0.91	103.36

Models	MAE ↓	RMSE ↓	MAPE (%) ↓	R^2 ↑	RT ↓
LSTM	4848	6355	0.061	0.90	133.09
MLP	3393	5156	0.042	0.91	115.47
SBEM	3381	5071	0.041	0.95	53.53

*Note: The values of significance are presented in bold font; "↓" signifies that a lower value is more favorable; "↑" indicates that a higher value is preferable.

4.4. Error Analysis

To visually contrast the average performance and stability of the six individual base prediction models and the proposed model, error analysis based on MAE and RMSE was conducted. The outcomes of each model, obtained from 10 separate runs, are depicted through box plots. The corresponding box plots are illustrated in Figure 2, which displays the distribution of these results.

As depicted in Figure 2, the interquartile range of the SBEM is narrower than that of the six individual base prediction models, which reveals its better stability. In addition, the average values of MAE and RMSE for the SBEM are lower than those of the base prediction models, demonstrating its superior average performance.

4.5. Analysis of Curve Fitting Results

To present the performance of each model in STLF in a more intuitive manner, Figure 3 illustrates the level of agreement between the actual values and the predicted values for each model.

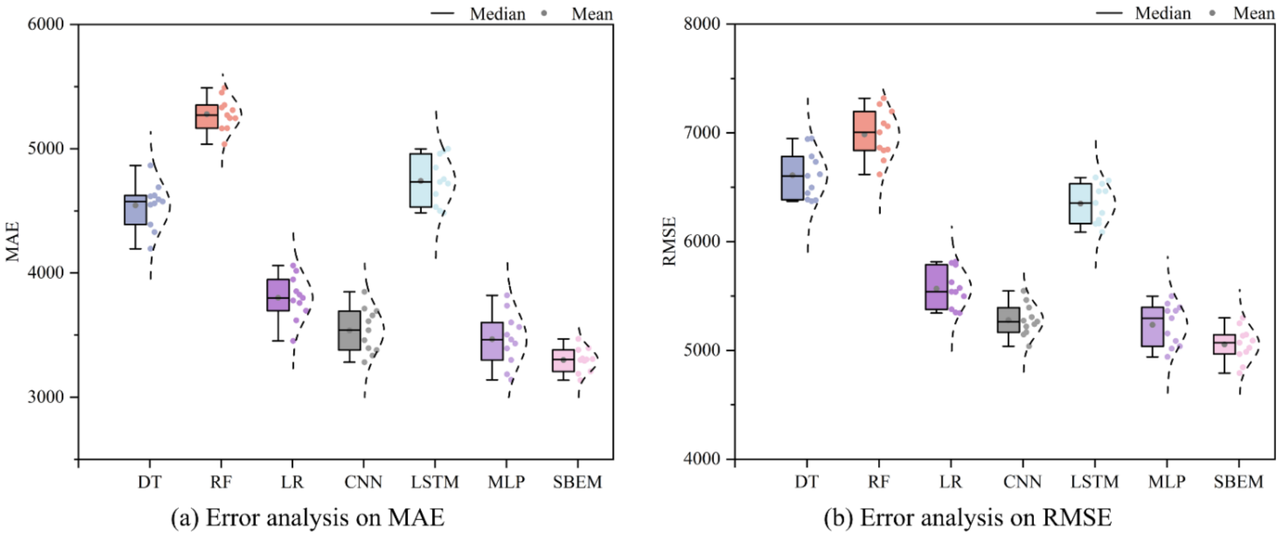


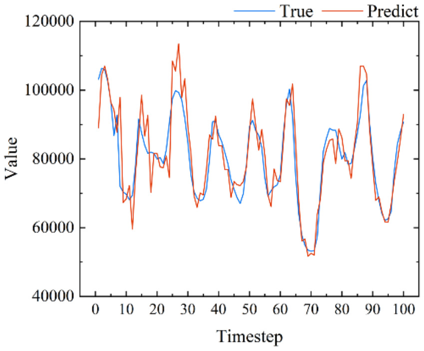
Figure 2. The structure of the SBEM.

As depicted in Figure 3, the predicted values derived from the SBEM are much nearer to the actual values. This suggests that when it comes to predicting electric load, the SBEM outperforms individual base prediction models. It is more precise and stable in grasping the variation trend of electric load.

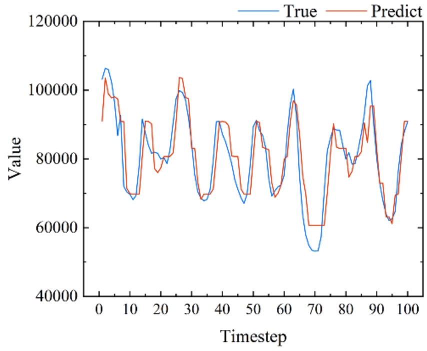
5. Conclusion

In this study, a stacking-based ensemble learning model is proposed for STLF. Firstly, min-max normalization and zero imputation are used for data preprocessing. Secondly, six base prediction models including LR, DT, RF, MLP, CNN, and LSTM are utilized to train optimized base prediction models. Thirdly, the optimized base prediction models are integrated to predict electric load with a stacking-based ensemble learning method. Experiments are extensively conducted to show the effectiveness of the SBEM.

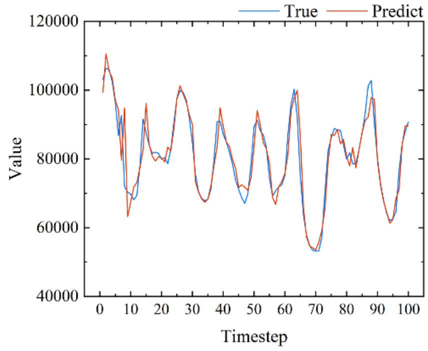
Besides the application in electric load forecasting, the proposed model has the potential for expansion into other prediction scenarios, such as wind prediction and electricity price prediction. In future studies, other relevant factors such as calendar indicators, natural disasters, and meteorological information will be incorporated. Besides, to enhance the interpretability of the proposed model, explainable machine learning will be investigated.



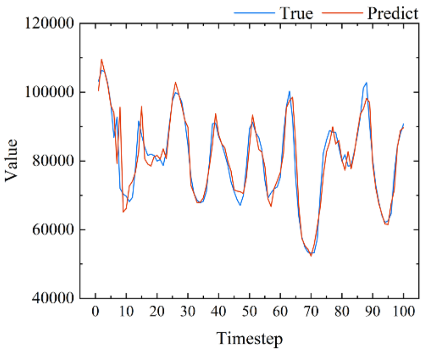
(a) Curve fitting results on DT



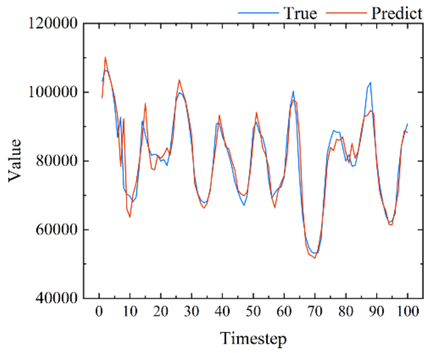
(b) Curve fitting results on RF



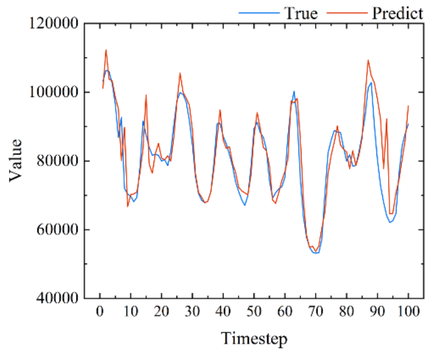
(c) Curve fitting results on LR



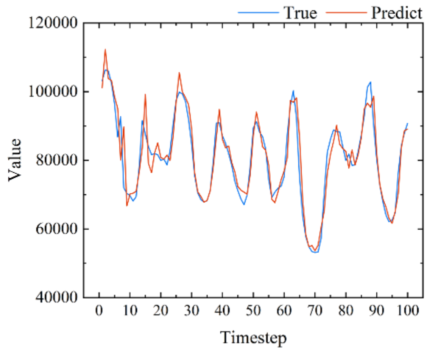
(d) Curve fitting results on CNN



(e) Curve fitting results on LSTM



(f) Curve fitting results on MLP



(g) Curve fitting results on SBEM

***Figure 3**. Curve fitting results.*

Abbreviations

STLF	Short-Term Load Forecasting
ARIMA	Autoregressive Integrated Moving Average
ES	Exponential Smoothing
MLR	Multiple Linear Regression
ANN	Artificial Neural Network
SVM	Support Vector Machine
RF	Random Forest
SBEM	Stacking-Based Ensemble Learning Model
LR	Logistic Regression
DT	Decision Tree
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Sankalpa, C., Kittipiyakul, S., & Laitrakun, S. (2022). Forecasting short-term electricity load using validated ensemble learning. *Energies*, 15(22), 8567. https://doi.org/10.3390/en15228567

[2] Wang, Y., Chen, Q. X., Hong, T., & Kang, C. Q. (2019). Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, 10(3), 3125-3148. https://doi.org/10.1109/TSG.2018.2818167

[3] Lee, C. M., & Ko, C. N. (2011). Short-term load forecasti ng using lifting scheme and ARIMA models. *Expert Syste ms with Applications*, 38(5), 5902-5911. https://doi.org/10.1016/j.eswa.2010.11.033

[4] Göb, R., Lurz, K., & Pievatolo, A. (2013). Electrical load forecasting by exponential smoothing with covariates. *Applied Stochastic Models in Business and Industry*, 29(6), 629-645. https://doi.org/10.1002/asmb.2008

[5] Fang, T. T., & Lahdelma, R. (2016). Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Applied energy*, 179, 544-552. https://doi.org/10.1016/j.apenergy.2016.06.133

[6] Feng, R. H., Xue, Y. B., Wang, W., & Xiao, M. (2022). Saturated load forecasting based on clustering and logistic iterative regression. *Electric Power Systems Research*, 202, 107604. https://doi.org/10.1016/j.epsr.2021.107604

[7] Hsu, C. C., & Chen, C. Y. (2003). Regional load forecasting in Taiwan—applications of artificial neural networks. *Energy conversion and Management*, 44(12), 1941-1949. https://doi.org/10.1016/S0196-8904(02)00225-X

[8] Dai, X. Q., Sheng, K. C., & Shu, F. Z. (2022). Ship power load forecasting based on PSO-SVM. *Mathematical Biosciences and Engineering*, 19, 4547-4567. https://doi.org/10.3934/mbe.2022210

[9] Fan, G. F., Peng, L. L., & Hong, W. C. (2022). Short-term load forecasting based on empirical wavelet transform and random forest. *Electrical Engineering*, 104(6), 4433-4449. https://doi.org/10.1007/s00202-022-01628-y

[10] Dietrich, B., Walther, J., Weigold, M., & Abele, E. (2020). Machine learning based very short term load forecasting o f machine tools. *Applied Energy*, 276, 115440. https://doi.org/10.1016/j.apenergy.2020.115440

[11] Liu, L. Y., Bai, F. F., Su, C. Y., Ma, C. P., Yan, R. F., Li, H. L., Sun, Q., & Wennersten, R. (2022). Forecasting the occurrence of extreme electricity prices using a multivariat e logistic regression model. *Energy*, 247, 123417. https://doi.org/10.1016/j.energy.2022.123417

[12] Ramos, D., Faria, P., Morais, A., & Vale, Z. (2022). Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. *Energy Reports*, 8, 417-422. https://doi.org/10.1016/j.egyr.2022.01.046

[13] Moon, J., Kim, Y., Son, M., & Hwang, E. (2018). Hybrid short-term load forecasting scheme using random forest an d multilayer perceptron. *Energies*, 11(12), 3283. https://doi.org/10.3390/en11123283

[14] Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 182, 72-81. https://doi.org/10.1016/j.energy.2019.05.230

[15] Kong, W. C., Dong, Z. Y., Jia, Y. W., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1), 841-851. https://doi.org/10.1109/TSG.2017.2753802

[16] Shi, J., Qu, X. L., & Zeng, S. T. (2011). Short-term wind power generation forecasting: Direct versus indirect ARIMA-based approaches. *International Journal of Green Energy*, 8(1), 100-112. https://doi.org/10.1080/15435075.2011.546755

[17] Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50, 1352-1372. https://doi.org/10.1016/j.rser.2015.04.065

[18] Xu, L., Wang, S. W., & Tang, R. (2019). Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Applied Energy*, 237, 180-195. https://doi.org/10.1016/j.apenergy.2019.01.022

[19] Qiu, Y. L., He, Z., Zhang, W. Y., Yin, X., & Ni, C. J. (2024). MSGCN-ISTL: A multi-scaled self-attention-enhanced graph convolutional network with improved STL decomposition for probabilistic load forecasting. *Expert Systems with Applications*, 238, 121737. https://doi.org/10.1016/j.eswa.2023.121737