JTE
HCMUTE
ISSN: 2615-9740

JOURNAL OF TECHNICAL EDUCATION SCIENCE
Ho Chi Minh City University of Technology and Education
Website: https://jte.edu.vn
Email: jte@hcmute.edu.vn

Enhanced Monthly Load Forecasting With RapidMiner-Based Deep Learning

Gia-Tue Tang, Dinh-Nhon Truong*
Ho Chi Minh City University of Technology and Education, Vietnam

*Corresponding author. *Email:* nhontd@hcmute.edu.vn

ARTICLE INFO

Received:	20/05/2025
Revised:	01/07/2025
Accepted:	25/07/2025
Published:	

KEYWORDS

Electric Load Forecasting;
Rapid Miner Model;
Artificial Intelligence;
Deep Learning;
Feature Selection.

ABSTRACT

Precise electrical demand forecasting is crucial for maintaining the reliability of the electricity supply, particularly in large urban centers. This study developed an artificial intelligence model with the ability to forecast daily electricity load over several months with high accuracy. The proposed model was trained and validated using historical energy consumption and meteorological data in a case study carried out in Ho Chi Minh City, Vietnam. Unlike previous MATLAB studies, this study employed the RapidMiner program to reduce calculation time and give a visual framework. The mean absolute percent error (MAPE) was used to evaluate prediction performance, yielding a MAPE of 0.52%, compared to 1.1% for Decision Tree and 8.9% for Support Vector Machine. Testing demonstrated that the proposed Deep Learning model significantly outperformed the baseline models. By incorporating feature extraction and explainability techniques, the model achieved high sensitivity to fluctuations, as indicated by an R-squared (R^2) value of 0.99. These results suggest that the model is practical for real-world applications and can assist in improving power system operation planning.

Doi: https://doi.org/10.54644/jte.2025.1901

1. Introduction

Load forecasting is crucial considering the magnitude of the growing demand for power brought on by industrialization and the advancement of electrical technology, especially for industrial and service-centric urban areas like Ho Chi Minh City. The energy demand increases annually due to significant fluctuations in peak-hour electricity consumption, weather-related fluctuations caused by the usage of heating and cooling equipment, and the expanding trend of electric vehicle charging stations in the system. Predicting energy load accurately reduces supply-demand imbalance occurrences, backup power generating costs, and enhances the reliability of electricity supply for residents and businesses. Therefore, in order to support grid management, a highly accurate and regularly updated power load forecasting model must be constructed.

Many approaches focus on short-term load forecasting (STLF) because longer horizons make it challenging to achieve the ideal forecast accuracy [1]. T. Hong discussed linear regression and artificial intelligence (AI) techniques before suggesting the potential of AI in time series data forecasting [2]. Recent AI approaches to the STLF problem are representative as Recurrent Neural Networks [3], Support Vector Machines [4], Long Short-Term Memory (LSTM) [5], Prophet algorithm [6], and hybrid approaches [7], [8], [9]. These efforts aim to lower prediction errors compared to conventional techniques and expand the prediction range. H. H. Goh *et al*. developed a combination of Multiple Convolutional Neural Networks with Long Short-Term Memory Network (MCNN-LSTM) model that attained a MAPE of 2.1% for STLF in a 24-step evaluation on the Ireland test dataset [10]. The model developed by M. Zhang *et al*. utilizing an input focus and an invisible connection mechanism built from RNN, can lower MAPE on the EirGrid dataset from 4.64% (using LSTM) to 0.52% in 121 seconds [11]. In Vietnam, the margin for electricity load forecasting is less than 2% [12]. Nguyen,T.H. *et al*. proposed a hybrid Harris Hawks Optimization–Wavenet (HHO–Wavenet) model for microgrid load forecasting on the Ho Chi Minh City power grid, achieving lower RMSE and MAPE than standard LSTM and CNN-based approaches [13]. In another study, Nguyen, T.H. *et al*. used Wavenet and Graph Convolutional Networks (GCN) in MATLAB attaining a MAPE of 1.34% in 1158 seconds, outperforming prior work [14]. However, these methods focus primarily on STLF, requiring significant computational resources and processing time, limiting their applicability to medium-term forecasting.

To bridge this gap, this study developed an artificial intelligence model using the RapidMiner tool [15] to forecast daily electricity load over a monthly period. The model incorporated prediction explainability techniques and input feature extraction to continuously refine the training process and reduce forecasting errors. This approach enabled the model to learn patterns, recognize trends, and self-adjust to dynamic system changes. The results demonstrate improved accuracy in medium-term load forecasting and enhanced adaptability for demand-side management applications.

2. Method for designing AI models

2.1. Preprocessing

Geographical load characteristics, socioeconomic factors, weather, and cultural events all influence electricity demand. The use of cooling equipment during the hot season in Ho Chi Minh City leads to high electricity consumption; therefore, weather conditions were considered in this study. The primary data source was the total electricity consumption of Ho Chi Minh City, recorded by smart meters at 30-minute intervals from January 2018 to March 2024. Due to the COVID-19 pandemic, the 2021 load data was unavailable. The peak load in the study was up to 94,847,172 kWh, and the lowest load value was 33,368,270 kWh. Weather information, including average, maximum, and minimum temperatures, sunrise and sunset times, precipitation, and wind speed, was collected from historical Meteo-Open data [16]. This dataset exhibits a complex and non-linear relationship between inputs and electricity consumption. Figure 1 illustrates the load diagram generated from the original input data.

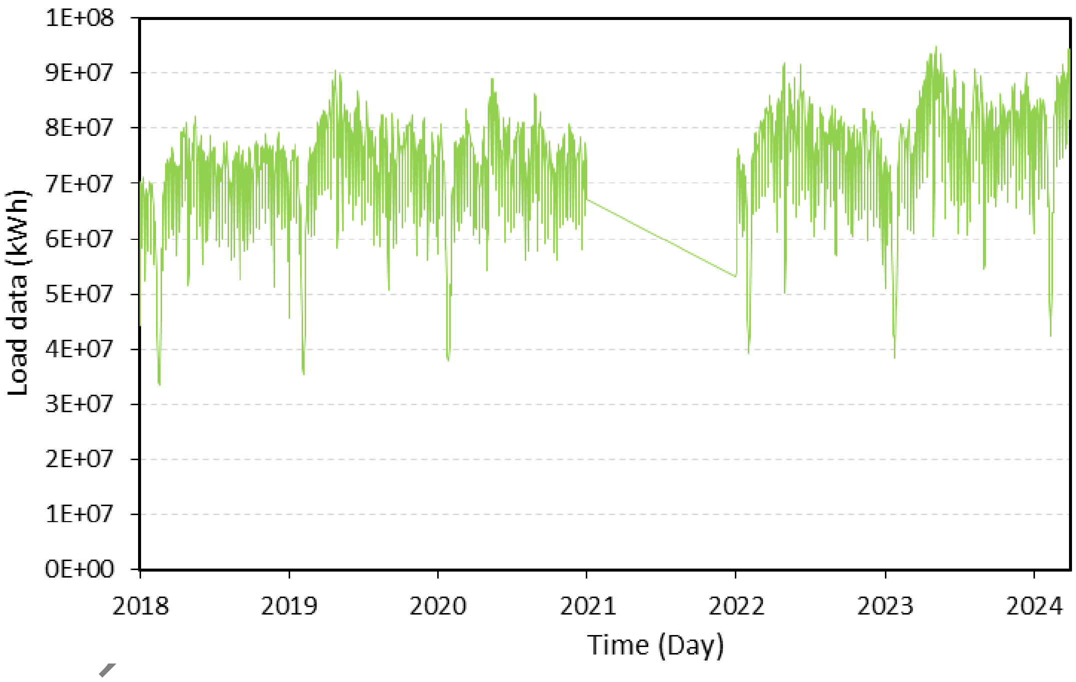

Figure 1. *Daily electricity consumption*

Missing or infinite values were identified and handled. Missing numeric values were imputed by the median of the surrounding points. The time attribute was transformed into a numeric feature by separating the time column into minutes, hours, days, months, and years. A sequence number column was added to help the model recognize growth trends. The power consumption attribute is labeled target. Feature labels are applied to the other properties. The data preprocessing process was illustrated in Figure 2. The most relevant features for the learning model were evaluated through preliminary testing on a sample dataset. The results section displays the percentage influence of the feature. Characteristics unhelpful to the learning process can be removed, or new features can be chosen for tracking the load variations. Seventy percent of the dataset is utilized for training, and thirty percent is used for testing.

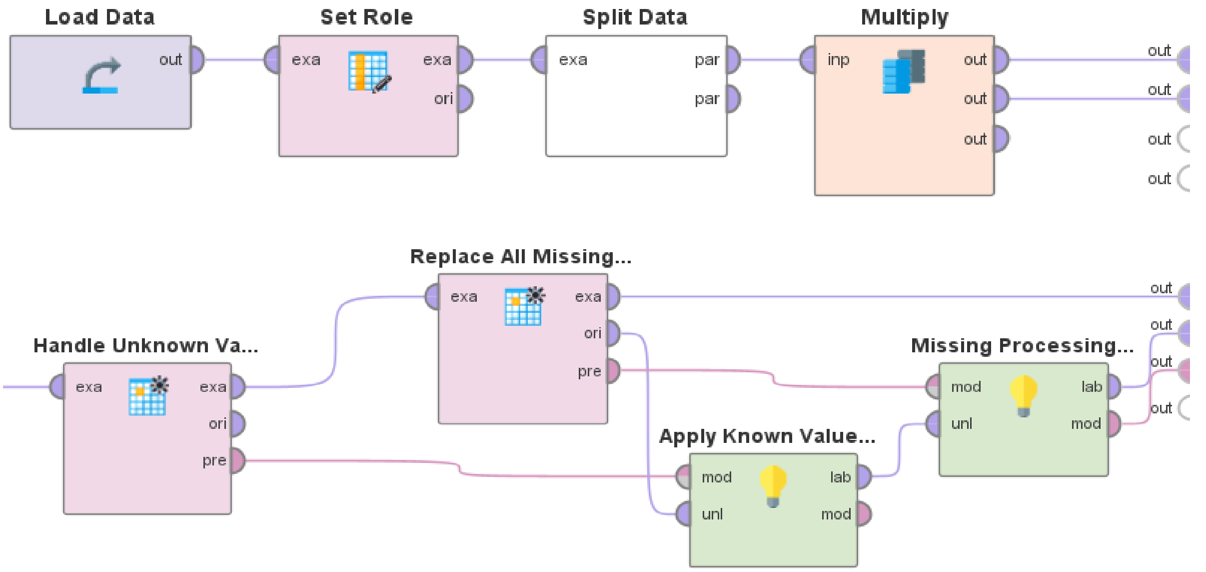

Figure 2. *Load Data*

By adding a phase to the pipeline that automated the entire input feature selection and optimization process, preprocessing time was reduced, redundant features were eliminated, and the accuracy of the resulting model was improved. A total of 2,500 samples were randomly selected from the dataset to evaluate which attributes produced the best results. The steps for feature extraction are illustrated in Figure 3.

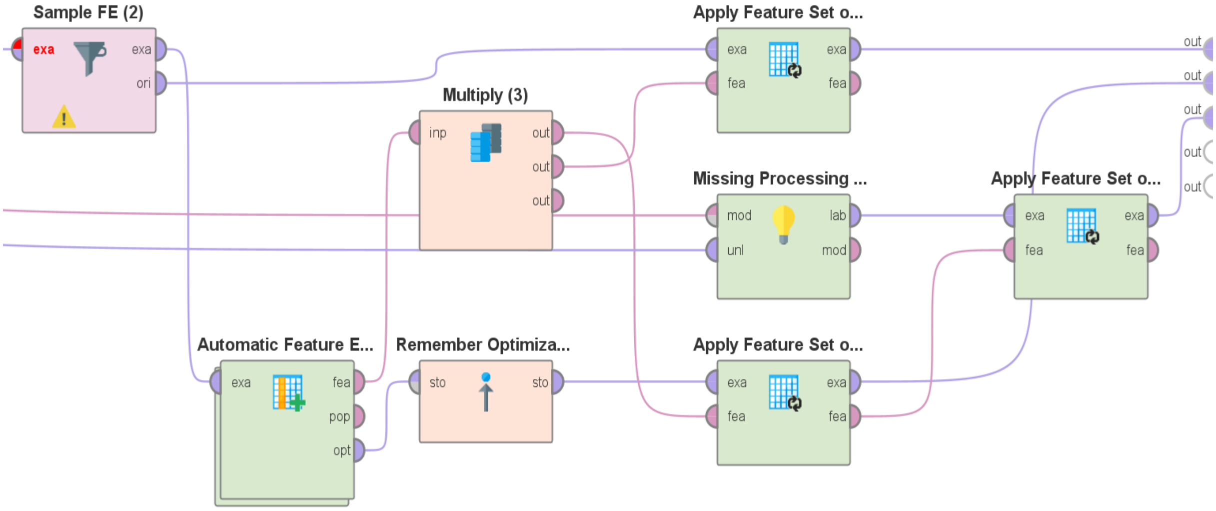

Figure 3. *Features Extract*

Figure 4 presents the workflow of the proposed forecasting model designed and implemented using RapidMiner. The process begins with data loading and preprocessing, including standardization and missing value handling. The dataset is then split into training and testing subsets. Feature extraction is performed to enhance the quality of learning before model training, evaluation, and optimization. Once the model achieves a relative error (RE) below 2%, it proceeds to the monitoring stage. If performance degradation is detected over time, retraining is automatically triggered to maintain accuracy. Upon final validation, the model outputs the forecasting results.

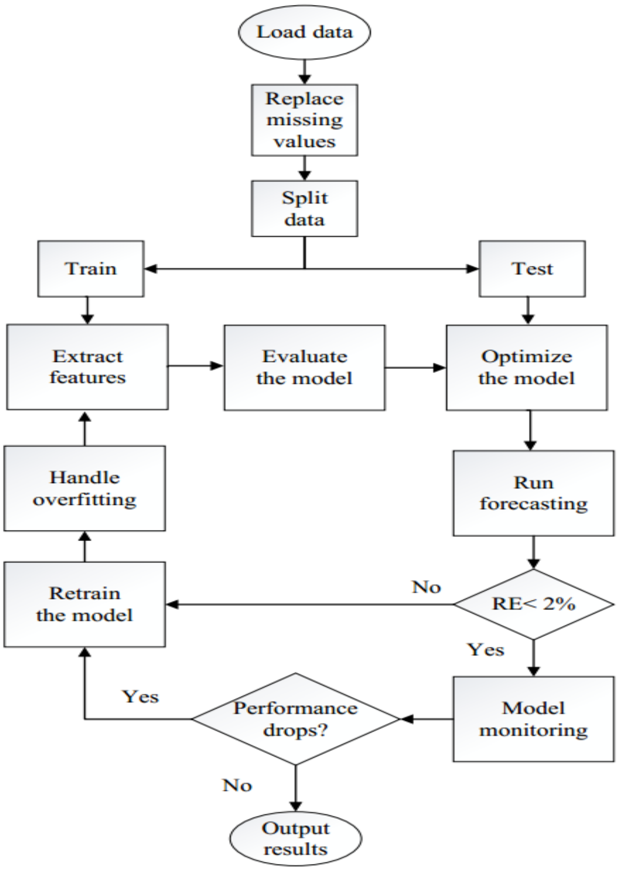

Figure 4. *Workflow of the proposed forecasting model developed using RapidMiner*

2.2. Training the model

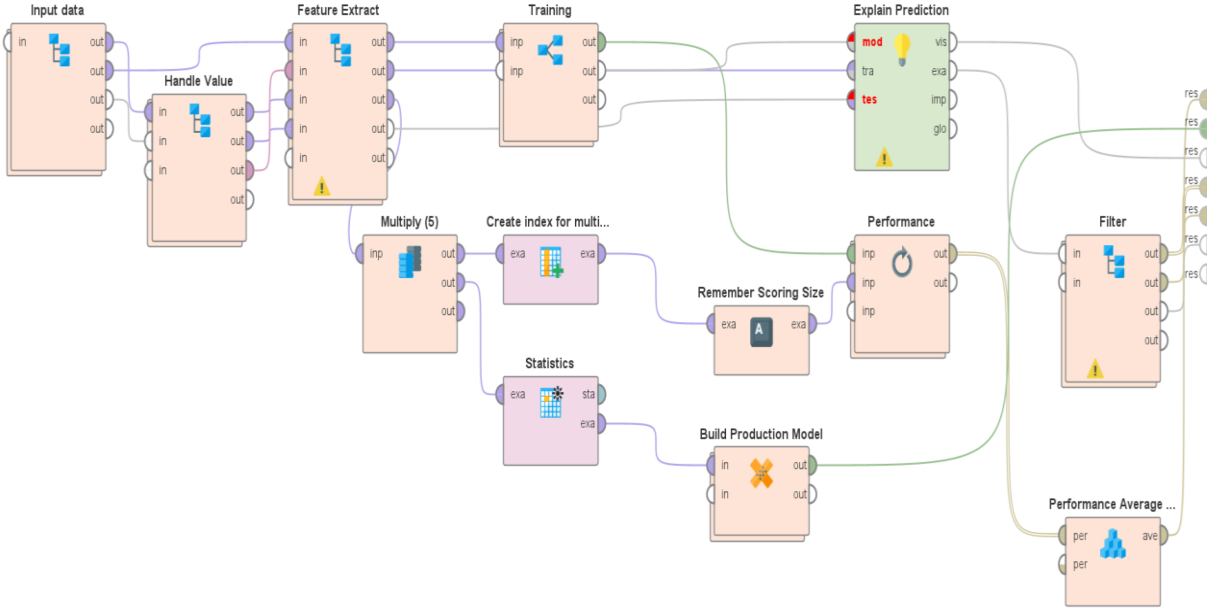

Figure 5. *Proposed AI Model*

The primary component of the model consists of Deep Learning Operators with a Multi-Layer Feed-Forward Artificial Neural Network (ML-FFNN) [15]. The basic idea of the algorithm is to simulate the behavior of neurons in the human brain. Figure 5 displays the proposed model.

The Deep Learning model employed in this study was trained on 11,500 data samples. The network architecture consists of two hidden layers, each containing 50 neurons. Both hidden layers utilized the

Rectifier (ReLU) activation function to capture non-linear relationships within the data. The output layer includes a single neuron corresponding to the target variable. To reduce training time, the number of epochs was set to 5. In order to mitigate overfitting, both L1 and L2 regularization techniques were applied with minimal values of 0.00001. The detailed configuration of the Deep Learning model is presented in Table 1.

Table 1. *Set value in Deep learning*

Rank	Parameter	Description	Set Value
1	Activation	Utilizing a mechanism for activation in hidden layers	Rectifier
2	The dimensions of hidden layers	The number of neurons and hidden layers it contains	2 layers 50 neurons per layer
3	Epochs	Number of times the entire dataset is passed through the network	5.0
4	Epsilon	Small value to avoid division by zero in calculations	1.0E-8
5	Rho	Parameter for adjusting the adaptive learning rate	0.95
6	L1	Parameter to prevent overfitting	0.00001
7	L2	Parameter to prevent overfitting	0.00001
8	Max w2	Limit on the squared weight values	10.0

As illustrated in Figure 6, the proposed model training process utilizes a ML-FFNN implemented via the H2O Deep Learning operator in RapidMiner. The model is trained using a loss function optimized through backpropagation and the gradient descent algorithm. Two deep learning blocks are employed for distinct purposes: one to extract hidden features from the input data, and another to learn from these features and perform prediction. The initial training block (Model PO) focuses on representation learning, while the second block (Model) uses these representations for accurate load forecasting. This architecture enhances predictive performance and demonstrates the effectiveness of deep learning in capturing nonlinear patterns in electricity load data.

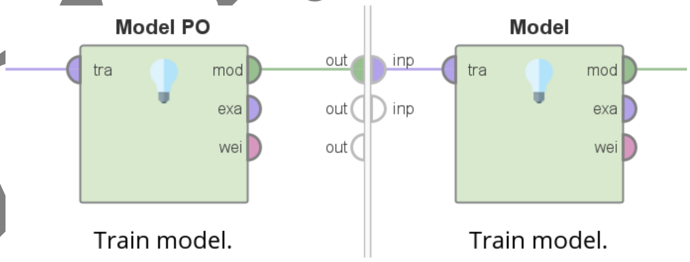

Figure 6. *Model Training Process Using Deep Learning*

2.3. Testing the model

How accurate the suggested model is verified based on the test dataset. The following formula is used to express the forecast error metrics. The more realistic the model, the smaller the forecast inaccuracy, and vice versa.

The square root of the overall forecast average divergence is known as the Root Mean Squared Error, or RMSE [1]. This value measures the overall error.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{i(pred)} - y_{i(act)}\right)^2} \tag{1}$$

Where, $y_{i(pred)}$ is the predicted daily electricity consumption and $y_{i(act)}$ is the electricity consumption according to historical data.

The average difference between the actual value and the anticipated value, expressed as a percentage of the actual value, is determined by the Mean Absolute Percentage Error (MAPE) [1].

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\left|y_{i(pred)} - y_{i(act)}\right|}{y_{i(act)}}\right)\times 100 \tag{2}$$

RapidMiner does not have an MAPE index; nevertheless, its counterpart, Relative Error (RE), can be applied to measure the ability to predict. An Arithmetic Operator was incorporated to compute the relative error for each data point. Therefore, the Relative Error result in this study can be considered MAPE.

$$RE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\left|y_{i(pred)} - y_{i(act)}\right|}{y_{i(act)}}\right) \tag{3}$$

R-squared (R^2) shows the percentage of actual value variance that can be accounted for by the regression model, where $\overline{y}$ is the average electricity consumption on day *i*.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{i(pred)} - y_{i(act)}\right)^2}{\sum_{i=1}^{N}\left(y_{i(act)} - \overline{y}\right)^2} \tag{4}$$

The predicted accuracy of the model gains with decreasing RMSE, MAPE, and RE error indices. The more precisely the model captures the variations in actual data, the closer the R^2 value is to 1.

2.4. Explain the Predicting

The proposed model was trained using a Deep Learning algorithm, where the internal weights significantly influenced the output [17]. In this context, the weights represent the importance of each input feature in contributing to energy consumption prediction. A weight value close to 1 indicates that the model has learned a strong and consistent association between that feature and the target output, which can directly affect the prediction accuracy.

Figure 7 provides an analysis of feature importance with respect to the model's prediction performance. This process helps identify which features are most valuable for forecasting and which may be considered noise or irrelevant. High-weight values were observed for specific time frames throughout the day, such as 12:30, 13:30, 16:00, 18:00, and 21:00, indicating peak electricity usage periods in Ho Chi Minh City, consistent with real-world consumption patterns.

Although features such as average daily temperature and sunrise and sunset times exhibited lower weights, they still contributed to the model's predictive capability, particularly in monthly and annual cycles. This behavior confirms that the model has successfully learned to associate temporal and environmental features with load patterns.

Following the interpretation of prediction results, the model can be further fine-tuned by adjusting the weights. A retraining loop may be applied until the Mean Absolute Percentage Error (MAPE) is reduced to 2% or less. While this step belongs to the data pretreatment phase, it also serves as an adaptive mechanism: if the prediction accuracy falls below 98%, the model will reprocess the input and continue learning until acceptable performance is achieved.

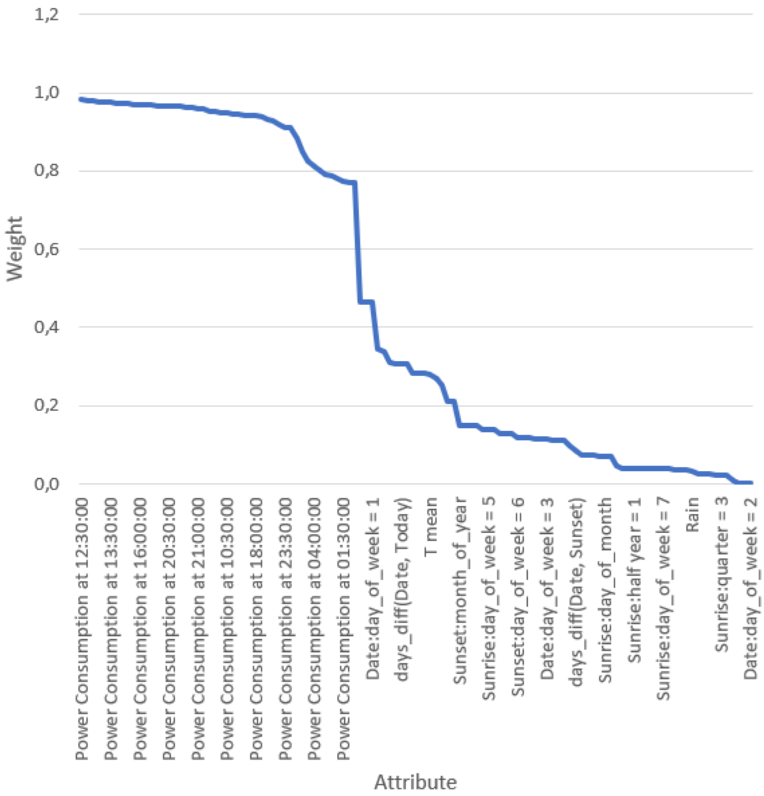

Figure 7. *Weight of Features*

3. Results

The model was evaluated using Altair AI Studio Engine 10.4.3 Educational Edition on an Acer Nitro AN515-57 laptop with standard hardware specifications (11th Gen Intel® Core™ i7-11800H and 8GB RAM). With the capabilities of the selected tools, computation times were reduced to 6 seconds for the Decision Tree model, 12 seconds for the Deep Learning model, and 69 seconds for the Support Vector Machine. This represents a notable improvement over prior studies conducted using MATLAB-based approaches in terms of processing efficiency. The proposed model demonstrated a strong ability to capture data fluctuations, as reflected by an R^2 value of 0.999. Furthermore, the model achieved high predictive accuracy without signs of overfitting or underfitting, evidenced by a remarkably low RE of 0.52%.

Table 2. *Comparison Result*

Rank	Model	RMSE (kWh)	RE (%)
1	Deep Learning	457,849.7	0.52%
2	Decision Tree	1,030,699.6	1.1%
3	Support Vector Machine	9,724,733.5	8.9%

Based on the performance comparison presented in Table 2, the Deep Learning model outperformed the other machine learning techniques evaluated in this study. The model demonstrated excellent accuracy and robustness in forecasting monthly electricity demand, achieving an RMSE of 457,849.7 kWh. In contrast, the Decision Tree and SVM models produced significantly higher RMSE values of 1,030,699.6 kWh and 9,724,733.5 kWh, respectively. Their RE was also substantially larger, with 1.1% for Decision Tree and 8.9% for SVM. These results confirm the superior predictive capability of the proposed Deep Learning approach.

Figure 8 shows a time series comparison between the actual and predicted daily electricity load in Ho Chi Minh City from February 2018 to March 2024. The green line represents the actual load, while the blue line shows the predicted values generated by the proposed Deep Learning model. The dotted red line represents the linear trend of the actual load. As observed, the predicted values closely follow the actual load curve throughout the entire period, including seasonal fluctuations and demand peaks. The alignment of the blue and green lines indicates that the model effectively captures both short-term variations and long-term trends in electricity consumption. Even during periods of high volatility, such as Tet holidays or summer peaks, the prediction remains accurate with minimal deviation. This strong fit confirms the model's ability to generalize well over an extended timeframe.

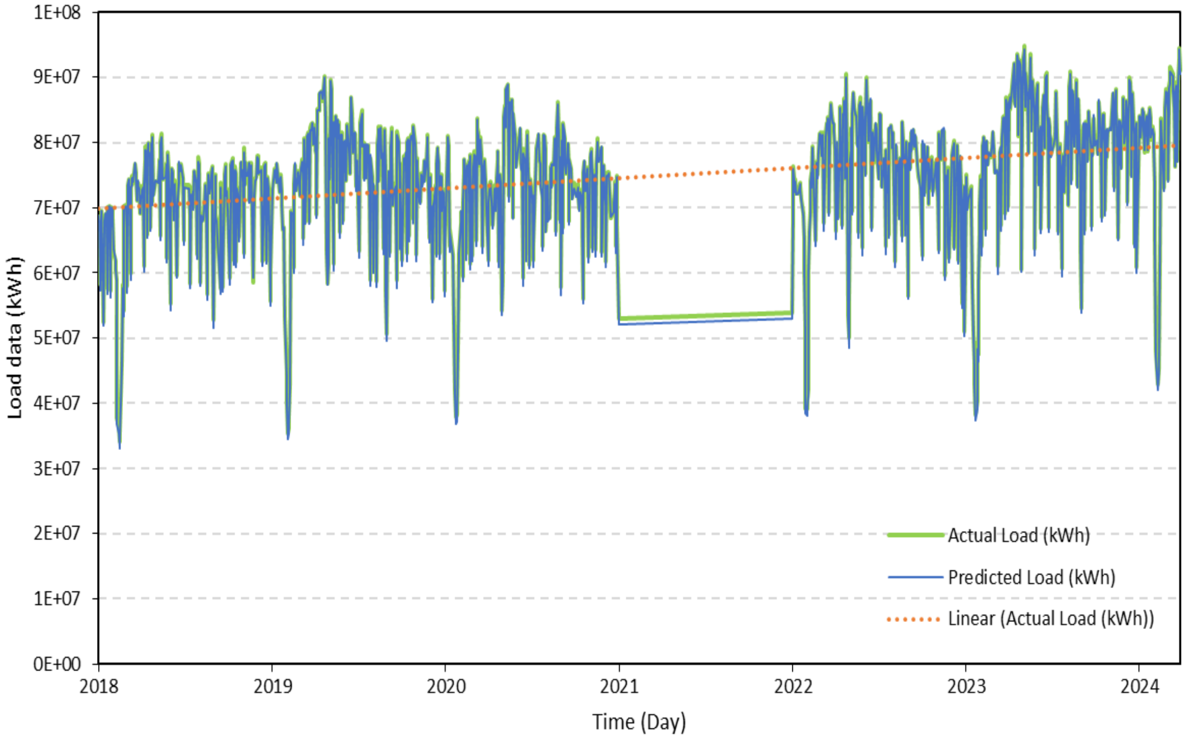

Figure 8. *A comparison between the actual and predicted daily load*

To investigate the impact of training duration on model performance, the Deep Learning model was trained with varying numbers of epochs: 3, 5, 10, 20, and 50. The evaluation was conducted based on two key performance metrics: RMSE and RE. The results are summarized in Table 3.

Table 3. *Effect of Increasing Epochs on RMSE and Relative Error in Deep Learning Model*

Rank	Epochs	RMSE (kWh)	RE (%)
1	3	820,060.7	0.94
2	5	457,849.7	0.52
3	10	457,849.7	0.52
4	20	249,024.5	0.26
5	50	249,024.5	0.26

With the hyperparameters configured in Table 2, the model achieved a RE of 0.52% after 5 epochs. In contrast, using only 3 epochs led to significantly higher errors (RE = 0.94%, RMSE = 820,060.7 kWh), indicating inadequate learning. As presented in Table 3, increasing the number of epochs notably improved performance, reducing RE by half and lowering RMSE to 249,024.5 kWh. However, improvements became marginal beyond 20 epochs, suggesting that the model had reached convergence. Throughout training, test errors remained low and the R^2 value consistently approached 1.0, indicating strong generalization and no signs of overfitting. Therefore, training with 50 epochs offers an effective balance between prediction accuracy and computational efficiency.

To support this learning, historical electricity and weather data from January 2018 to March 2024 were used. Early data (2018–2020) helped the model recognize long-term trends, including pre- and post-pandemic variations. Results are presented for January 2023 to March 2024 to reflect current conditions and enable detailed monthly-level evaluation.

Table 4. *Monthly Forecasting Performance Metrics: RMSE and RE from January 2023 to March 2024*

Rank	Month	RMSE (kWh)	RE (%)
1	Jan-23	684,373.0	1.11
2	Feb-23	599,947.5	0.81
3	Mar-23	568,426.5	0.71
4	Apr-23	518,971.1	0.62
5	May-23	494,433.3	0.57
6	Jun-23	475,359.4	0.53
7	Jul-23	476,963.2	0.52
8	Aug-23	462,451.8	0.5
9	Sep-23	476,400.1	0.52
10	Oct-23	465,082.1	0.5
11	Nov-23	453,912.2	0.48
12	Dec-23	446,870.6	0.47
13	Jan-24	295838.6	0.26
14	Feb-24	419,391.6	0.48
15	Mar-24	407,677.3	0.45

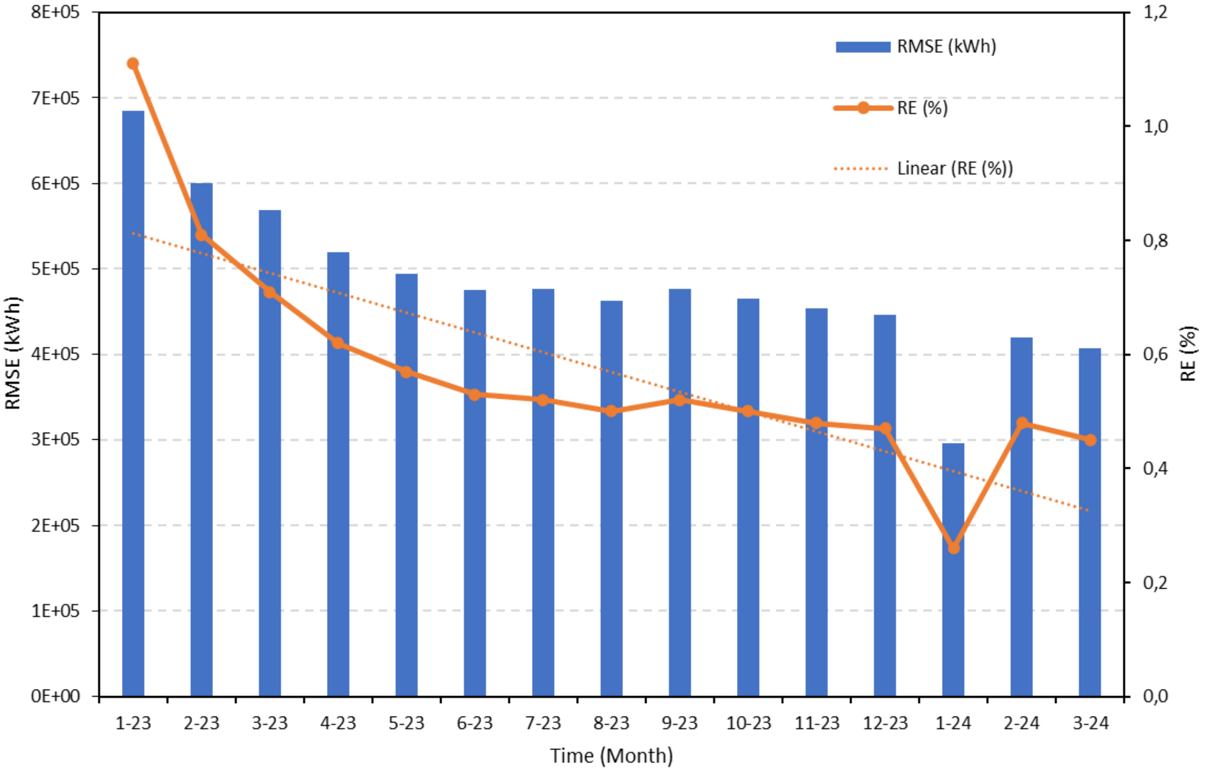

Figure 9. *Monthly RMSE and RE of the proposed model from January 2023 to March 2024*

As shown in Table 4 and Figure 9, the forecast results reflect realistic seasonal variations in Ho Chi Minh City. Higher errors in early 2023 (RE = 1.11%, RMSE = 684,373 kWh in January) likely stem from the Tet holiday, which drives unpredictable spikes in electricity usage. In contrast, from May to December 2023, RE values consistently stayed below 0.6%, indicating that the model effectively learned seasonal patterns—capturing increased loads in the dry months (March–May) and steadier demand during the rainy season (June–October).

Forecast performance improved further in early 2024, with the lowest RE values recorded in January and February (0.26% and 0.48%, respectively), supported by the model's exposure to extensive, multi-year data. Accuracy also remained strong in transitional months like March and October, suggesting robustness to seasonal shifts.

Overall, the declining trend in both RMSE and RE confirms the model's growing predictive strength. Its consistent performance across various consumption phases demonstrates suitability for real-world deployment in power system planning and demand-side management.

4. Conclusions

The findings of this study indicate that the proposed Deep Learning approach, implemented on the RapidMiner platform, offers an efficient and robust solution for forecasting monthly electricity load in Ho Chi Minh City. Compared to MATLAB-based approaches, RapidMiner significantly reduces the programming effort and improves model prototyping, making it accessible for broader application in energy informatics. The model demonstrated strong adaptability to local consumption patterns by automatically extracting relevant features and producing accurate predictions. However, the current setup does not incorporate external socio-economic and policy-related variables. Factors such as electricity pricing schemes, household income levels, and the adoption rate of electric vehicles may strongly influence future demand but were not included due to data limitations. To address this and enhance both the accuracy and practical applicability of load forecasting, we propose integrating advanced monitoring and control solutions into the RapidMiner framework. Specifically, real-time data collection via the IoT system [18] will support continuous forecasting model calibration, while the energy storage control strategy [19] helps optimize the response to load fluctuations, ensuring grid stability and efficiency. This combination aims to build a synchronous, flexible energy management system, reduce power loss and support the development of smart grids in the future.

Acknowledgments

This work is part of the author's Master's thesis conducted within the Faculty of Electrical and Electronics Engineering at Ho Chi Minh City University of Technology and Education. The author gratefully acknowledges the University for providing access to laboratory facilities, equipment, and a supportive academic environment.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. A. Rob and J. Hyndman, Eds., *Forecasting: Principles and Practice*, 23rd–25 Sept ed., University of Western Australia.

[2] T. Hong, "Energy forecasting: past, present and future," *Foresight: Int. J. Appl. Forecasting*, no. 32, pp. 43–48, 2014.

[3] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018, doi: 10.1109/TSG.2017.2686012.

[4] R. A. Khan, C. L. Dewangan, S. C. Srivastava, and S. Chakrabarti, "Short term load forecasting using SVM models," in *Proc. 8th IEEE Power India Int. Conf. (PIICON)*, Jul. 2018, doi: 10.1109/POWERI.2018.8704366.

[5] Z. Zhang, X. Wang, J. Wang, and X. Yuan, "Power load forecasting model based on LSTM and Prophet algorithm," in *Proc. 3rd Int. Conf. Consumer Electron. Comput. Eng. (ICCECE)*, pp. 156–159, 2023, doi: 10.1109/ICCECE58074.2023.10135304.

[6] X. Pei, "Prophet algorithm-based power load forecasting model," in *Proc. IEEE 3rd Int. Conf. Electron. Technol., Commun. Inf. (ICETCI)*, pp. 1498–1503, 2023, doi: 10.1109/ICETCI57876.2023.10176778.

[7] Y. Wang, N. Zhang, and X. Chen, "A short-term residential load forecasting model based on LSTM recurrent neural network considering weather features," *Energies*, vol. 14, no. 10, p. 2737, May 2021, doi: 10.3390/en14102737.

[8] A. Mathew *et al.*, “Medium-term feeder load forecasting and boosting peak accuracy prediction using the PWP-XGBoost model,” *Electr. Power Syst. Res.*, vol. 237, p. 111051, Dec. 2024, doi: 10.1016/j.epsr.2024.111051.

[9] Ş. Özdemir, Y. Demir, and Ö. Yildirim, “The effect of input length on prediction accuracy in short-term multi-step electricity load forecasting: A CNN-LSTM approach,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3540636.

[10] H. H. Goh *et al.*, “Multi-convolution feature extraction and recurrent neural network dependent model for short-term load forecasting,” *IEEE Access*, vol. 9, pp. 118528–118540, 2021, doi: 10.1109/ACCESS.2021.3107954.

[11] M. Zhang, Z. Yu, and Z. Xu, “Short-term load forecasting using recurrent neural networks with input attention mechanism and hidden connection mechanism,” *IEEE Access*, vol. 8, pp. 186514–186529, 2020, doi: 10.1109/ACCESS.2020.3029224.

[12] Ministry of Industry and Trade, *Vietnam Electricity Development Plan VIII*, 2023. (in Vietnamese)

[13] T. H. Nguyen *et al.*, “Hybrid HHO-Wavenet model applies in short-term load forecasting for microgrid system,” in *Proc. 6th Int. Conf. Green Technol. Sustainable Dev. (GTSD)*, pp. 1177–1183, 2022, doi: 10.1109/GTSD54989.2022.9989027.

[14] T. H. Nguyen *et al.*, “Application of GCN-Wavenet hybrid model in short-term load forecasting for microgrid systems,” (in Vietnamese), *J. Sci. Technol. – Univ. Danang*, vol. 20, no. 11.2, 2022. [Online]. Available: https://jst-ud.vn/jst-ud/article/view/8031

[15] “Deep Learning - Altair RapidMiner Documentation.” [Online]. Available: https://docs.rapidminer.com

[16] “Historical Weather API | Open-Meteo.com.” [Online]. Available: https://open-meteo.com

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[18] D. N. Truong, M. S. N. Thi, V. T. Ngo, and A. Q. Hoang, “Development of the monitoring program for an integrated small-scale wind and solar systems based on IoT technology,” *J. Sci. Technol.: Inf. Commun. Technol.*, vol. 19, no. 12.2, pp. 26–31, Dec. 2021, doi: 10.31130/ICT-UD.2021.139.

[19] H. T. Pham, D. N. Truong, V. P. Vu, and H. A. Trinh, “Design of Battery Control Strategy for Hybrid PV/BESS System,” (in Vietnamese), *J. Tech. Educ. Sci.*, vol. 20, no. 02(V), pp. 77–88, May 2025, doi: 10.54644/jte.2025.1805.

Gia-Tue Tang graduated from the Industrial University of Ho Chi Minh City, Vietnam, in 2023 with a B. Eng. degree in Electrical and Electronic Engineering. She is currently pursuing an M.Eng. degree in Electrical Engineering at Ho Chi Minh City University of Technology and Education. Her areas of research interest are power system stability management.

Email: 2430615@student.hcmute.edu.vn. ORCID: https://orcid.org/0009-0009-2992-3259

Dinh-Nhon Truong, Ph.D., is an Associate Professor at the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Education (HCMUTE), Vietnam. He earned his Ph.D. in Electrical Engineering from National Cheng Kung University (NCKU), Tainan, Taiwan. His research interests encompass power system stability, renewable energy conversion systems, microgrids, and FACTS devices.

Email: nhontd@hcmute.edu.vn. ORCID: https://orcid.org/0000-0002-4015-6769