Enhanced Data-Driven Load Forecasting Framework for High PV Penetration Electrical System: A Case Study of SUT's Campus

Nisachon Thabcha[1], **Nitikorn Junhuathon**[2], and **Keerati Chayakulkheeree**[1†], Non-members

ABSTRACT

This research proposes an approach to develop a one-day-ahead electrical load forecasting model using deep learning techniques. It compares deep learning models with basic models, namely Moving Average, DNN, LSTM, Bi-LSTM, CNN-LSTM, and Attention-LSTM, to improve accuracy and reliability through Mutual Information (MI) and Shapley Additive Explanation (SHAP) analysis. This approach utilizes data collection and feature engineering to optimize the context of an electrical loads with integrated solar power generation, Using Suranaree University of Technology (SUT) campus electrical system as a case study. Three accuracy indices are examined: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). Among the NN-based models, Bi-LSTM and LSTM tend to provide the best overall forecasting performance from the input data, and the Bi-LSTM model is the most accurate model with the lowest metric value among all models. The lag load feature, or the value of the electrical load in the past one day, is the feature that is most related to the forecasting target and has the most impact on the model's decision.

Keywords: Load Forecasting, Long Short-Term Memory, Bidirectional Long Short-Term Memory, Mutual Information, SHAP, Explainable AI

1. INTRODUCTION

It is well known that electric power is one of the most widely used energy sources and has been used for a long time, both in people's daily lives and well-being and as an important part of national and economic development [1]. The greater the economic growth, the greater the population's demand for electricity [2]. Electric power must, therefore, be reliable and managed with quality [3]. In Thailand, there has been a continuous increase in renewable energy and the introduction of electric vehicles, resulting in unpredictable patterns of electricity usage, with fluctuations that are difficult to manage [4]. Therefore, developing accurate and transparent electricity load forecasting is important for managing electricity to achieve maximum efficiency. It is also linked to the National Energy Plan (NEP2024), which defines five pillars of electric power as follows: Pillar 1: Demand Response & Energy Management System (DR & EMS). Load forecasting is the basic information that the EMS system uses to plan production and load management in advance. This includes coordinating the load response for maximum efficiency, reducing peak demand, and increasing the stability of the power system. Pillar 2: Renewable Energy Forecast (RE Forecast), load forecasting and renewable energy are in tandem in maintaining the power system balance, especially when renewable energy is highly volatile. Accurate load forecasting will reduce the uncertainty in power dispatch and support the adoption of a higher proportion of renewable energy into the system. Pillar 3: Microgrid & Prosumer Local Energy Systems, load forecasting in this context enables the appropriate management of energy resources in each area, reducing the risk of power outages or overloads. Pillar 4: Energy Storage System (ESS) and Pillar 5: EV Integration, although not directly related, load forecasting helps to manage charging and energy storage or discharge behaviour more efficiently. If energy storage systems (batteries) are installed in the future or electric vehicles (EVs) are used at a high level, load forecasting will play a role in designing systems to support charging and energy usage at different times [5].

There are three popular methods for forecasting electricity load. Regression-based methods are considered simple models, such as linear regression, which work well with statistically recorded data and determine the relationship between inputs and outputs. However, complex and non-linear energy demand data used for forecasting can pose problems. Time series methods, typically used with time-dependent data, such as ARIMA models, face challenges when load changes suddenly and due to external factors, making the original data difficult to predict and uncertain [6]. Finally, artificial intelligence (AI) and deep learning (DL) techniques are used. These models are more adaptable to large, voluminous, volatile, and complex data sets. Models belonging to the Neural

Manuscript received on October 6, 2025; revised on January 1, 2026; accepted on January 5, 2026. This paper was recommended by Associate Editor Chawasak Rakpenthai.

[1]The authors are with School of Electrical Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand.

[2]The author is with Department of Electrical Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand.

[†]Corresponding author: keerati.ch@sut.ac.th

Digital Object Identifier: 10.37936/ecti-eec.2026241.261546

Network (NN) family are currently trending towards widespread use in forecasting [7]. Table 1 presents a summary of key issues of each research study, including feature engineering, selecting the best model from a comparison of forecasts with other models, metrics used for evaluation, techniques used for feature selection, and model explainability, to address gaps in research and the objectives of this study.

However, it has been found that many works employ deep learning models in the context of load forecasting. Deep learning models are more effective at learning temporal relationships and non-linear behavior than traditional load forecasting methods, coupled with the selection and design of contextual feature engineering.

However, there is a lack of comparisons between diverse models, including traditional methods and deep learning or NN models, including feedforward models compared to RNNs, models that incorporate RNN models, and hybrid models. Furthermore, there is a gap in predicting features that are effective for nonlinear data, clearly explaining the results of black-box models, and identifying features that truly influence the model's priorities. The trend in load forecasting model development in research focuses on interpretability to enhance both efficiency and transparency. Many studies have focused on using SHAP and other Explainable Artificial Intelligence (XAI) techniques to analyze feature importance, such that the results obtained after model training are designed to directly assess feature importance. However, this lack of integration with methods that filter data before modeling is found. Popular MI methods select accurate features before modeling to reduce model bias. Furthermore, some studies have found that using attention weights can only describe the timesteps that the model focuses on for model explanation. Therefore, our research aims to develop approaches to address these challenges. This leads to the following main objectives:

1) To develop a one-day electricity load forecasting model using deep learning techniques. We will compare deep learning models with basic models, including Moving Average, DNN, LSTM, Bi-LSTM, CNN-LSTM, and Attention-LSTM. Specifically, we will examine the model with the highest accuracy based on real-world load data reconstructed from electricity meters and solar panels.

2) To improve accuracy and reliability through the design of appropriate and high-quality features. Mutual Information (MI) is a promising approach for capturing complex nonlinear energy consumption patterns. and systematic evaluation using performance indicators including RMSE, MAE, and MAPE.

3) To improve the transparency and interpretability of the model, SHAP analysis is used to identify influential features and analyze their significance. All features in the dataset can be described to provide a deeper understanding of the model's performance.

This research focuses on the accuracy and reliability of one-day-ahead electricity load forecasting at Suranaree University of Technology (SUT). The project uses real-world load data from the Provincial Electricity Authority (PEA), collected every 15 minutes for a year. Deep learning techniques using artificial neural networks compare models from various categories. Data is prepared, and feature engineering is developed based on the context of the area to reflect the university's actual energy demand. The importance of features can be explained both before and after input into the model for electricity load forecasting. The highlight of this work lies in the systematic design of the forecasting process to achieve both model accuracy and feature importance analysis through the integration of MI and SHAP methods. Most studies use only one technique for feature evaluation and evaluating features with non-NN models due to their inherent black-box nature. However, this work interprets features from deep learning models and conducts two-step feature analysis. The first step, using MI, analyzes the relationship between real data and load, providing a statistical overview. The second step, using SHAP, analyzes the impact on actual model decisions, demonstrating the model's performance on the data, comparing and utilizing the advantages of both methods for optimal forecasting performance. Furthermore, black-box models can be explained and data analyzed before being imported into the model.

At the institutional level, this model provides a more accurate energy planning approach, enabling demand management and cost optimization, as well as more flexible power system deployment [18]. It also employs a data-driven organization that can support peak demand management [19]. The impact on the energy sector extends beyond university energy management and can be scaled up to other contexts, including smart grid energy management. This approach improves energy management efficiency across various contexts by accurately forecasting energy demand. Energy suppliers can plan energy production and distribution, reduce overload, and effectively reduce energy losses in transmission lines through integration with energy management technologies in smart grids [20]. In the industrial energy management sector, it helps forecast energy demand for equipment or processes, helping energy managers plan when to use energy (demand-side management), reducing electricity costs during periods of high energy prices, and making production systems more sustainable [21]. In the conjunction with the renewable energy sources sector, accurate forecasting models can be used to forecast energy demand in conjunction with renewable energy sources, which can help reduce energy imbalances and improve the efficiency of renewable energy use [22]. In the Greenhouse Gas Emissions sector, it supports reducing unnecessary energy use as part of a strategy to reduce greenhouse gas emissions in the energy sector [23]. This paper is divided into four sections: Section 1 introduces the study and stating the research objectives; Section 2 presents the research methodology used; Section 3 presents the study

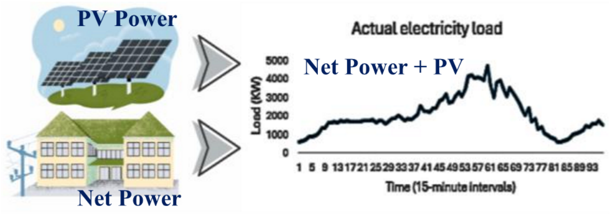

Fig. 1: *Preparation of Actual Electricity Load Data.*

results and discussion; and Section 4 concludes with a conclusion.

2. METHODOLOGY

2.1 Data Collection and Preparation

Before entering the model for load forecasting, Suranaree University of Technology's electricity consumption data was collected for 1 year, collected every 15 minutes in 2023, totalling 35,040 data sets, which is considered high-resolution data for load forecasting. The data was collected from Automatic Metering (AMR) maintained by the Provincial Electricity Authority (PEA). In addition, solar energy generation data was collected from an Automatic Metering (AMR) maintained by an external company, ENOS, with the same resolution as the electricity load data. Weather data and academic calendar data for that year were collected and prepared to input all this data into the future feature engineering method. The required data is the actual load data. Therefore, the load data obtained from the meter was combined with the solar energy generation data to obtain the data showing the actual load of the university, as shown in Fig. 1.

The preparation of the real electricity load data can be shown on Fig.1. All data used as features and output for the model to learn were divided into two parts: the first part was the data set to be trained, and the other part was partitioned for model validation. Time Series Cross-Validation (TSCV) was performed on this data set, which is a method of data partitioning to evaluate models that take into account temporal order and prevent future data leakage. The training and validation sets were generated using a Time Series Split with 5 folds. The second part was the data set tested on the model after training and validation were completed. It was a data set that the model had never seen before, so the results were unbiased and divided into 4608 data points. For testing, this group of data will be divided into 10 sets; each set will have 3 days. Each set is divided according to the types of data groups, namely long holiday, weekdays during the school term, exam period, summer season, the start of the semester, compensation holiday period, rainy season, weekdays and weekends, term break, and winter season.

2.2 Feature Engineering

In this research, feature engineering has a group of data that has the characteristics of feature creation, which is creating new features from existing time data and converting time data that may not be easy for the model to logical data with contextual meaning. It consists of the following techniques: In the feature transformation method, features that use this method include Weekdays and Weekend, Daytime and Nighttime, Compensatory Holiday, and Long Weekend, which are converting to variable values and binary representation with numbers 0 and 1. In the time-based feature engineering method, features that use this method include yearly hours, which divides the data into meaningful time sequences with 1 year's data every 15 minutes, creating features that allow the model to learn the day and time. The Time Component Extraction method features that use this method include Quarter-Hourly Interval, which extracts time components into separate features, creating new features that describe the sub-patterns that occur in each hour and the lag features method. Features that use this method include Lag Load, which creates features from previous values by taking historical load values to create features. In addition, there is a group of characteristics: external features. External variables affect the forecast of the target variable, such as ambient temperature, an external factor that directly affects energy use behavior, and dew point, an indicator of air humidity.

2.3 Feature Selection and Mutual Information

Mutual Information (MI) captures the relationship of input features to a non-linear data model, shows their importance, and indicates how the feature reduces uncertainty with respect to the forecasting target. It is used as a tool for selecting appropriate forecasting features, especially in tasks where the number of features varies without necessarily being linear. It is calculated using the following Eq. (1).

$$i\,(X;Y) = \sum_{x\in X}\sum_{y\in Y} p\,(x,y)\cdot\log\left(\frac{p(x,y)}{p(x)\cdot p(y)}\right), \tag{1}$$

where Y is the target variable, i.e., the future load, X is the input variable, $p(x, y)$ is the joint probability of x and y, and $p(x)$ and $p(y)$ are the marginal probabilities of each variable. The value of MI is expressed as a real number greater than or equal to 0. The higher the value of MI, the more the feature is related to the forecast value. MI can help reduce the size of the data to control both the speed of training and the interpretation of the model. MI is a powerful tool for feature selection for time series models, especially in complex load forecasting tasks that change over time and in different contexts. Selecting appropriate features using MI can make the model more accurate and easier to interpret.

Feature selection in load forecasting each feature plays a physical or behavioural role in load patterns. The Lag Load feature was selected because electrical loads are similar and have daily activities that often repeat within 24 hours, making it a very important representative of daily patterns in forecasting. The Yearly

Table 1: *Review the literature and summarize the objectives.*

Ref	Best Deep Learning Model	Feature Selection	Model Explainability	RMSE	MAE	MAPE
[4]	LSTM	MI	-	✓	✓	-
[8]	LSTM, Bi-LSTM	-	-	✓	✓	✓
[9]	-	-	SHAP value based	-	-	✓
[10]	RAID	-	SHAP	-	-	✓
[11]	-	Correlation	KAN	✓	-	-
[12]	LSTM-Attention-LSTM	Attention weights	Attention weights	✓	-	✓
[13]	Attention-LSTM	Attention weights	Attention weights	✓	✓	-
[14]	CNN-LSTM	Correlation	-	✓	✓	✓
[15]	Bi-LSTM	MI	-	✓	-	✓
[16]	ANN	MI	-	-	-	✓
[17]	MLP	-	SHAP	✓	-	✓
Proposed Framework	**6 models**	**MI**	**SHAP**	✓	✓	✓

Hours and Quarter-Hourly Interval features help the model understand the time sequence of each day and allow the model to recognize period differences. When used in combination with other features, the model can better learn about daily trends. The Weekdays and Weekend features were selected because electricity usage by electricity consumers differs significantly during holidays and weekdays. The Daytime and Nighttime features were selected because people's behaviour shows higher electricity usage in the morning than at night, so the model should be able to see this difference. The Compensatory Holiday and Long Weekend features have different load patterns during long holidays and special holidays. The Ambient Temperature and Dew Point features are weather-related features that reflect the characteristics of electricity consumers when temperatures and humidity change. They also provide the model with sufficient forecasting capabilities when there is load variation caused by changing weather conditions.

2.4 Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) model is a structure of Recurrent Neural Network (RNN) designed to solve problems and manipulate time-series data. It consists of a Forget Gate that determines which data should be forgotten from the original cell state. Let x_t be the input feature vector at time t, $\sigma(\cdot)$ be the sigmoid activation function, h_{t-1} be the hidden state from time t-1, and b_f be the bias vector of the forget gate, W_f be the learnable weight matrix of the forget gate. The forget-gate output is computed as shown in Eq. (2).

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right), \tag{2}$$

where $f_t \in [0,1]$ is close to 1, indicating retention, while values close to 0 indicate erasure. This helps determine whether data should be retained. The next component, the Input Gate (i_t) and the Candidate Cell State, decides which new data to add to the cell state, as shown in Eqs.

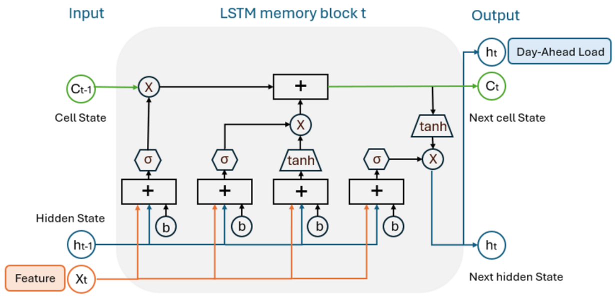

Fig. 2: *The LSTM Architecture.*

(3) and (4).

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \tag{3}$$

$$\tilde{C}_t = \sigma\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right), \tag{4}$$

where i_t controls which new data should be imported, and $\tilde{C}_t$ represents the new data processed (candidate memory). The next component, the Update Cell State, combines the Forget Gate and the Input Gate to create a new cell state, as shown in Eq. (5).

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{5}$$

where the forgotten data is multiplied by f_t, and the new, important data is multiplied by i_t and added together. The final section is the Output Gate (o_t) and Hidden State. These define the data to output as hidden states for future forecasting, as shown in Eqs. (6) and (7).

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right), \tag{6}$$

$$h_t = o_t * \tanh\left(C_t\right), \tag{7}$$

where h_t is the hidden state, used to predict the electrical load at the next point, as shown in Fig. 2.

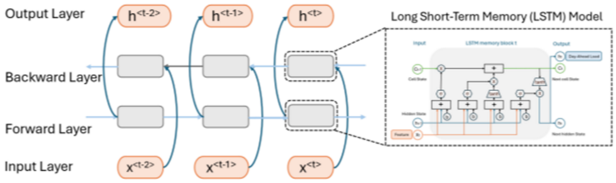

***Fig. 3**: The Bi-LSTM architecture.*

2.5 Bidirectional Long Short-Term Memory (Bi-LSTM)

The Bidirectional Long Short-Term Memory (Bi-LSTM) model is a structure that is developed from LSTM to increase the ability to learn sequences of data that have relationships in both the past (past dependencies) and the future (future dependencies) simultaneously. It consists of two LSTM layers that work together: a forward LSTM that processes data from the past to the present and a backward LSTM that processes data from the future back to the present at any time t. The results of both directions are combined into a single hidden state as shown in Eq. (8).

$$h_t = \left[\overrightarrow{h}_t; \overleftarrow{h}_t\right], \tag{8}$$

where $\overrightarrow{h}_t$ is the hidden state from the Forward LSTM, $\overleftarrow{h}_t$ is the hidden state from the Backward LSTM, and h_t is the combination of the two results, as shown in Fig. 3.

2.6 Shapley Additive Explanation (SHAP)

SHAP is a method based on the Shapley Value from Cooperative Game Theory, which measures how each feature contributes to the model's decision. SHAP transforms the model description problem into a game in which each feature is a player, and the model's outcome is a reward that must be shared (prediction). The Shapley Value calculation considers all subsets of features (feature coalitions) to assess the impact of each feature. The calculation principle (conceptually) can be described as the Shapley value of feature i is defined by the following Eq. (9).

$$\varnothing i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \cdot [f(S \cup \{i\}) - f(S)], \tag{9}$$

where N is the set of all features, S is the subset of features excluding i, $f(S)$ is the predicted value of the model when using only the feature in S, and $\varnothing i$ is the average of the effects that feature i has on the predicted value when combined with all other features in various orders.

2.7 Load Forecasting Framework

Electrical load forecasting requires the preparation of essential data, including aggregation and collection from reliable sources in the forecasting area, focusing on actual electricity load data from meters. This data is then fed into feature engineering. Two types of data are used: feature generation, which converts raw data into numbers that are easier for the model to understand and allows the model to better learn about data differences, and external feature data, which learns about data trends. This process takes place during the data collection and preparation process.

The next step in the simulation process is automatic feature selection using MI to assess the nonlinear relationship between each input variable and the target load. This helps preserve features that are important for forecasting. This reduces data complexity while retaining the most important features. The data is then grouped into three groups for training, validation, and testing, based on proportions. The resulting data is then fed into the model by designing the model, setting various parameters, and training deep learning models in various categories: Moving Average, DNN, LSTM, Bi-LSTM, CNN-LSTM, and Attention-LSTM. After training the models on the same dataset for all models, then, in the results and analysis process, the performance is evaluated using standard metrics: root means square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). To enhance the transparency and clarity of the black-box model, the SHAP tool is used to guide decision-making on feature prioritization. This tool identifies the lowest and highest priority features for the final pre-load forecast. Figure 4 shows the entire forecasting process. The first step is the data collection and preparation process. The next step is the simulation process. The last step is the results and analysis process.

***Table 2**: Performance indicators.*

Name of Criteria	Formula
Mean Average Percentage Error (MAPE	$\frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \hat{Y}_i)}{Y_i}$
Mean Absolute Error (MAE)	$\frac{\sum_{i=1}^{n}\lvert Y_i - \hat{Y}_i\rvert}{n}$
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$
Average (Avg.)	$\frac{\sum_{i=1}^{m} X_i}{m}$

2.8 Performance Factor

To examine the performance of the proposed forecasting model through evaluation indicators, namely percentage of MAPE, MAE and RMSE as shown in Table 2, the calculation of the average of each indicator is given by the equation of Average (Avg.), where X is the MAPE, MAE or RMSE indicator on day i and m is the total number of days. Let Y_i denote the actual load at time i, and $\hat{Y}_i$ denote the corresponding predicted load.

3. RESULTS AND DISCUSSION

The electrical load forecast from all the data obtained and the tools mentioned above will be used in simulation

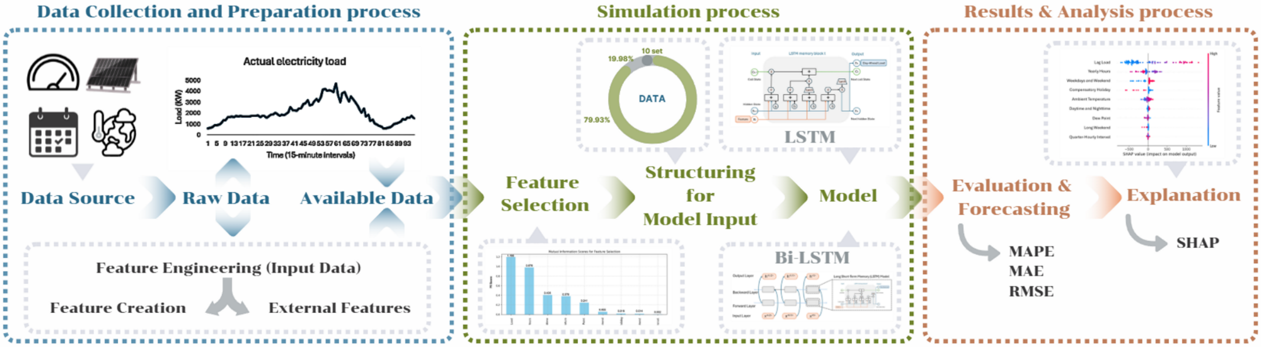

Fig. 4: The Complete forecasting process.

using the Python 3 program to show the results comprehensively and explain the accuracy and reliability of the model in the area and context of Suranaree University of Technology. It is a 1-day-ahead load forecast using historical data for 1 year of forecasting in 2023. The details are presented in this section.

The settings for each model are shown in Table 3. All models use similar parameters, namely the Dropout Rate parameter with a value of 0.2 to set the percentage of that unit. To prevent overfitting in the model, the learning rate is set to 0.001 to determine the speed of parameter tuning during learning. The optimizer uses Adam to automatically adjust the learning rate. The batch size, which is the number of data used to calculate the gradient per iteration of parameter tuning, is set to 32. The Epochs, which is the number of iterations to learn the whole data set, is set to 50. The loss function is used to evaluate the model error with MSE. In the different parts, each model is a layer part, which is the basic structure of the model that defines data processing. The DNN model has a Flatten Layer to transform the data and feed it into a Dense Layer, with the number of units in each Dense Layer having the units or the complexity of learning in each layer as 128, 64, and 1. The CNN-LSTM model has Conv1D filters, or the number of filters, as 64. The kernel size is the length of the repeated pattern detection filter with the value of 3. The pool size for reducing the dimensionality is set to 2. The LSTM Units are set to 100 for learning the temporal relationship after Conv1D has extracted features and has a Dense value of 1 in the model part. The Attention-LSTM has an Attention Layer that prioritizes the timesteps that affect the prediction and returns 100 LSTM Units with a Dense of 1. In the LSTM model, the LSTM Unit Capacity is 100, and the Bi-LSTM model has a Bidirectional Wrapper to process two directions and uses 100 LSTM Units in each direction with a Dense of 1.

The selection of the appropriate hyperparameter values for each model was performed using a trial-and-error method. The parameters were adjusted within a specified range and then evaluated by evaluating the minimum MAPE value displayed when each parameter was adjusted within each range. Hyperparameter units were set in a range of 64 to 128, with a best value of 100. The dropout rate was set in a range of 0.1 to 0.3, with the best value of 0.2. The learning rate was set in a range of 0.005 to 0.0005, with a best value of 0.001. Timesteps were set in a range of 4 to 12, with a best value of 8. Batch size was set in a range of 16 to 64, with a best value of 32. Early stopping was set in a range of 5 to 20, with a best value of 10. As shown in Table 4, these hyperparameter sets were then used to simulate the results of each model.

To prioritize the feature data before importing the model with the MI tool, the relationship of the features and the target electrical load dip data groups is measured as both linear and non-linear data. It shows the score ranking of each feature. The Lag Load feature has an MI score of 1.198, the Yearly Hours feature has an MI score of 0.976, the Daytime and Nighttime feature has an MI score of 0.405, the Ambient Temperature feature has an MI score of 0.378, the Dew Point feature has an MI score of 0.241, the Weekdays and Weekend feature has an MI score of 0.055, the Compensatory Holiday feature has an MI score of 0.018, the Long Weekend feature has an MI score of 0.014, and the Quarter-Hourly Interval feature has an MI score of 0.002, as shown in Fig. 5. The introduction of lag load as a feature tends to have a significant relationship between the input and output data. The characteristics of the Lag Load feature are most clearly related to the forecast value. The behaviour of electricity users today and yesterday is similar. The features that are tested with high scores are the Yearly Hours feature, which shows the time position that repeats patterns throughout the year, showing the relationship with the forecast value in a high proportion. These are 2 features that have significantly higher scores than other models. As for other features, the scores are closer to zero and are less relevant to the forecast.

One-day-ahead electrical load forecasting at Suranaree University of Technology can demonstrate the performance of each of the six models through three metrics, compiled into statistical metrics that show the highest, lowest, and average values of each metric from ten test data sets. Important days in the forecast, including the days showing the lowest and highest values of each metric, are also included in the test data set for the model. It is shown in Table 5. The forecasting results show that the moving average forecasting method gives the highest

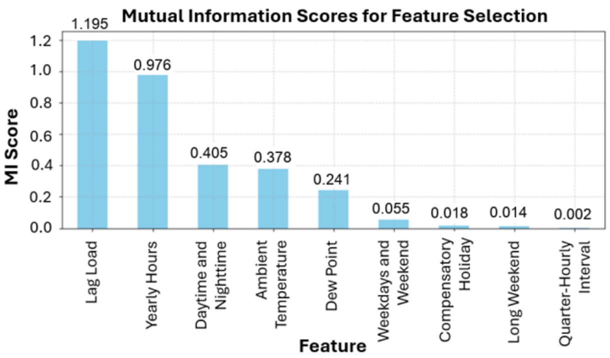

***Fig. 5**: Mutual information scores.*

***Table 3**: Parameter setting.*

Model	Parameters Layer
DNN	Flatten, Dense (128), Dense (64), Dense (1), and Units: 128, 64, 1
CNN-LSTM	Conv1D (64), kernel size (3), MaxPooling1D (pool size=2), Dense (1) and LSTM (100)
Attention-LSTM	LSTM (100), return sequences (True), Dense (1) and Attention
LSTM	LSTM (100) and Dense (1)
Bi-LSTM	Bidirectional (LSTM(100)) and Dense(1)

***Table 4**: Selecting the appropriate hyperparameter.*

Hyperparameter	Adjustment range	Best parameter values
Units	[64 128]	100
Dropout Rate	[0.1 0.3]	0.2
Learning Rate	[0.0005 0.005]	0.001
Timesteps	[4 12]	8
Batch Size	[16 64]	32
Early Stopping	[5 20]	10

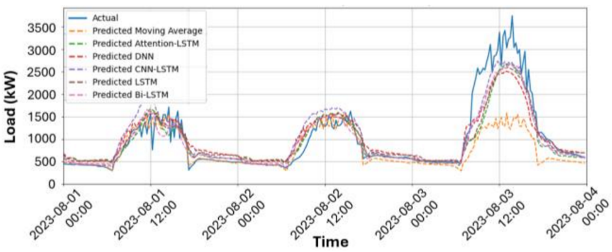

***Fig. 6**: The lowest-performing forecast results in the initial semester test set.*

value of all 3 indicators from the tested data set. MAE has a value of 675.72 kW, and RMSE has a value of 945.33 kW on 2023-08-03, which is a date in the set of semesters start dates, as shown in Fig. 6. And MAPE has a value of 78.34% on 2023-09-10, which is a date in the rainy season set, as shown in Fig. 7. The forecasting method with the Bi-LSTM model gives the lowest values of MAE, RMSE,

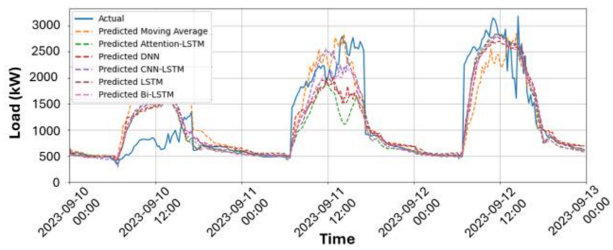

***Fig. 7**: The lowest-performing forecast results in the rainy season test set.*

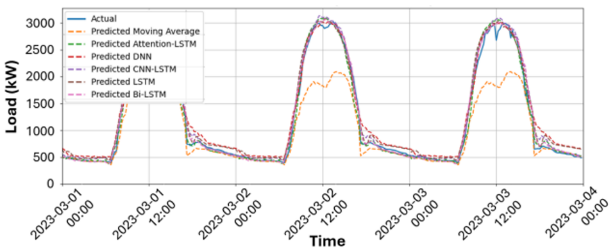

***Fig. 8**: The best-performing forecasting results in the test set during the exam period.*

and MAPE indicators from the tested data set, with values of 46.75 kW, 68.35 kW, and 4.82%, respectively, which are a date in the exam set, as shown in Fig. 8. The figures show the load forecast for 1 day in advance for 3 consecutive days. Each model is graphed compared to the actual load graph to show the forecast trend.

Since the moving average model relies on the average of historical load patterns, they must be similar at the same time point for accurate forecasting. However, data from the beginning of the semester, the rainy season, or the test data used for forecasting are volatile and unstable. This results in higher forecast error rates than all other models when analyzed for the maximum possible forecast error. The Bi-LSTM model has two layers that process historical data and predict future load trends, enabling it to effectively learn from data patterns that fluctuate with weather and time. Furthermore, the test data during exams or similar data have lag load and yearly hours features that significantly affect the relationship with the forecast at that time, accounting for a significant portion of the MI score. This factor occurs both before and after the forecast point. Therefore, the bidirectional learning method demonstrates the highest forecasting performance compared to all other models when analyzed for the maximum possible model accuracy. The Bi-LSTM model exhibits a lower average error (MAE) value, a percentage error compared to the actual value (MAPE), and a lower proportion of large errors (RMSE) than other models. The model can reduce errors at each time point, reduce forecasts with small errors, and forecast peak load periods more accurately than other models. Therefore, the resulting model is the best-performing model in this context.

The LSTM model showed the lowest values of all

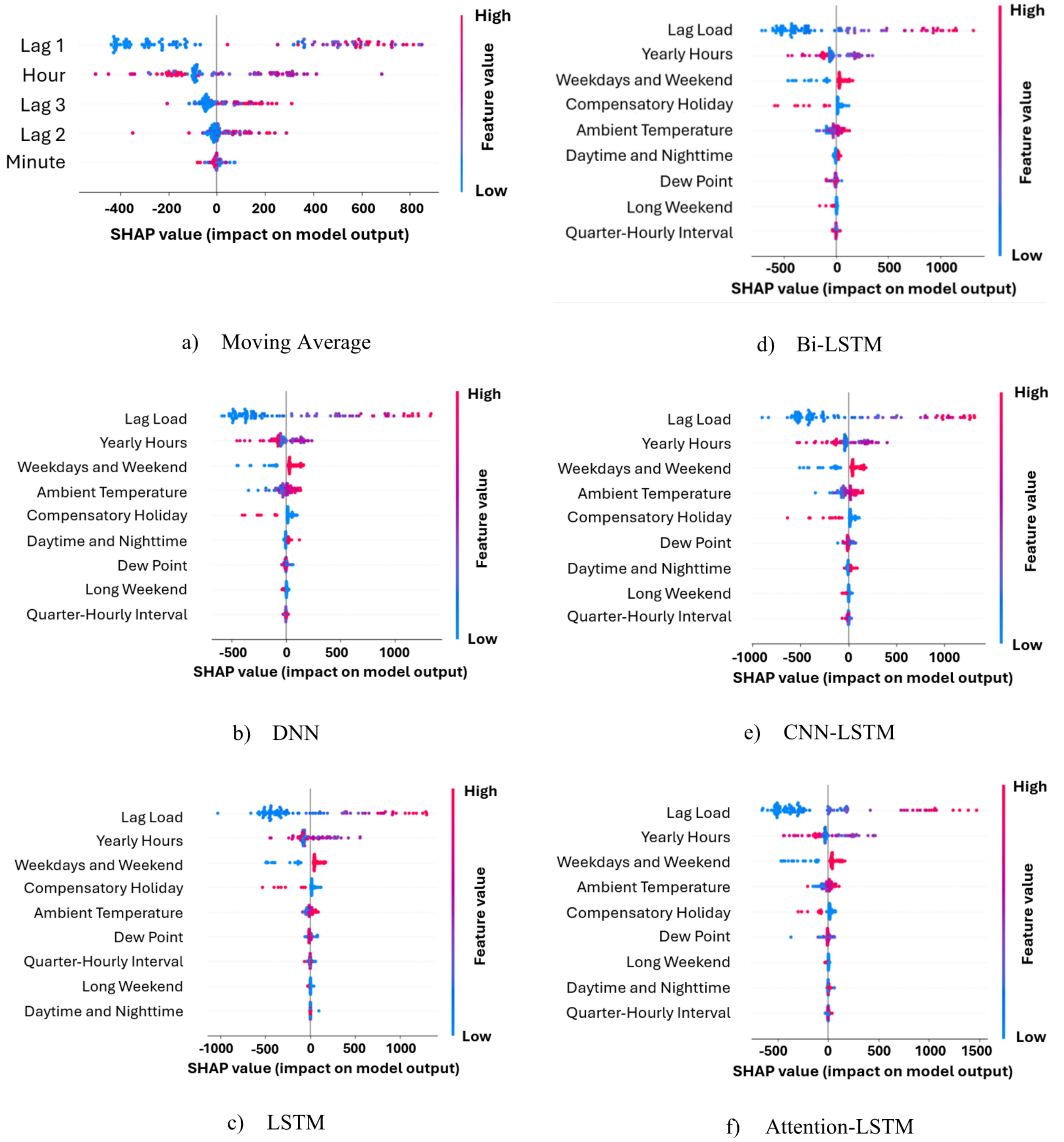

***Fig. 9**: Shap Value of Each Model: a) Moving average, b) DNN, c) LSTM, d) Bi-LSTM, e) CNN-LSTM, and f) Attention-LSTM.*

three indicators as follows: MAE was 83.60 kW, RMSE was 105.69 kW, and MAPE was 9.82%. The model has a learning characteristic from the past to the present. There is a layer that remembers various values, which is different from the DNN model in that the features fed into the model are not chronologically ordered, which makes the indicator values higher than the LSTM model. MAE was 97.47 kW, RMSE was 122.82 kW, and MAPE was 10.16%. For the hybrid CNN-LSTM model, CNN was used to capture short-term data patterns, resulting in lower errors. The lowest values of all three indicators were as follows: MAE was 68.33 kW, RMSE was 90.55 kW, and MAPE was 7.21%. For the Attention-LSTM model, developed from a simple LSTM structure, it prioritizes the most effective timestep, showing an MAE of 73.10 kW, an RMSE of 114.97 kW, and an MAPE of 5.99%. Both models exhibit higher forecasting accuracy than the simple LSTM model. The moving average method yields the best indicators, with an MAE of 68.60 kW, an RMSE of 87.39 kW, and an MAPE of 5.78%. These are more effective than some of the models mentioned above. However, when examining the worst-day recordings, the

***Table 5**: Performance of each model on 10 test sets.*

Models	Metric Statistic			
	Metric	Min.	Max.	Avg.
Moving average	MAE (kW)	68.60	675.72	292.63
	RMSE (kW)	87.39	945.33	461.50
	MAPE (%)	5.78	78.34	18.03
Attention-LSTM	MAE (kW)	73.10	385.39	192.96
	RMSE (kW)	114.97	563.92	299.34
	MAPE (%)	5.99	35.40	14.64
DNN	MAE (kW)	97.47	403.80	208.02
	RMSE (kW)	122.82	615.58	308.85
	MAPE (%)	10.16	37.38	17.03
CNN-LSTM	MAE (kW)	68.33	368.25	188.39
	RMSE (kW)	90.55	556.91	293.62
	MAPE (%)	7.21	39.63	14.28
LSTM	MAE (kW)	83.60	362.97	192.74
	RMSE (kW)	105.69	528.20	282.47
	MAPE (%)	9.82	33.84	15.87
Bi-LSTM	MAE (kW)	46.75	373.37	178.63
	RMSE (kW)	68.35	566.79	288.43
	MAPE (%)	4.82	38.74	12.90
Bi-LSTM (Top 3 Features by SHAP)	MAE (kW)	56.95	270.42	183.34
	RMSE (kW)	85.51	388.98	283.21
	MAPE (%)	4.89	36.79	13.99

model achieves higher indicators than all other models. Furthermore, the average test data also has the lowest overall performance.

The average of the indicators from all 10 test data sets, each set having 3 days or 30 data points, can be shown in order of the models with the lowest to highest average MAE indicators as follows: Bi-LSTM, CNN-LSTM, LSTM, Attention-LSTM, DNN, and moving average models with values of 178.63 kW, 188.39 kW, 192.74 kW, 192.96 kW, 208.02 kW, and 292.63 kW, respectively. The order of the models with the lowest to highest average RMSE indicators is as follows: LSTM, Bi-LSTM, CNN-LSTM, Attention-LSTM, DNN, and moving average models with values of 282.47 kW, 288.43 kW, 293.62 kW, 299.34 kW, 308.85 kW, and 461.50 kW, respectively. The model rankings with the lowest to highest average MAPE metrics are shown as Bi-LSTM, CNN-LSTM, Attention-LSTM, LSTM, DNN, and moving average, with values of 12.90%, 14.28%, 14.64%, 15.87%, 17.03%, and 18.03%, respectively, showing the prediction trend of each model from all the input test data. The Bi-LSTM and LSTM models still have good prediction performance overall.

In addition, the Bi-LSTM model was investigated to obtain the most three influential features from SHAP: Lag Load, Yearly Hours, Weekdays, and Weekend. The lowest values of the metrics were obtained as follows: MAE was 56.95 kW, RMSE was 85.51 kW, and MAPE was 4.89%. Compared to the same model set using all 9 features, the results show that the model using only the most important features provides acceptable forecast performance, although its performance is slightly reduced compared to using all features. The Lag Load feature has the highest weight in describing the daily electrical load characteristics. The Yearly Hours, Weekdays, and Weekend features still help the model better understand hourly and weekly patterns. While selecting only 3 important features allows the model to perform well, using all features yields the lowest metric values and the best performance.

The SHAP tool was imported to describe the importance of features that affect the decision of the moving average, DNN, LSTM, Bi-LSTM, CNN-LSTM, and Attention-LSTM models, respectively, as shown in Fig. 9. The x-axis shows the SHAP value, with a positive value indicating a higher prediction value. The y-axis sorts the features by cluster importance from highest to lowest, and the color of the dots indicates the actual feature value, with red having a high value and blue having a low value. The top three features of the NN model with the highest importance are as follows: The Lag Load feature is an important variable that has a great impact on the model's decision because it has the highest SHAP value distribution. There are a lot of red dots to the right, meaning that if yesterday's load is higher, it will also push today's forecast value higher. The Yearly Hours feature is distributed in clusters and has most red dots to the left, meaning that features with high values will lower the forecast value. The Weekdays and Weekend features show that the closer the SHAP value distribution is to zero, the less the feature's impact on the model's decision. In the non-linear features, the SHAP values are widely distributed, with both positive and negative values and a mixture of red and blue dots at the same point. The moving average method is a basic statistical method for forecasting using features Lag 1, Lag 2, and Lag 3 as the electrical load values for the past 1, 2, and 3 days, respectively. The time-related data, namely the Hour and Minute features, are the time data for all 3 days in every hour and minute at the same time of each day. Feature Lag 1 is the feature with the most dispersion and has the most impact on the model's decision.

The importance of ranking of features of NN models by SHAP and MI, where model A is LSTM, B is Bi-LSTM, C is CNN-LSTM, D is Attention-LSTM, E is DNN, and F is MI. From Table 6, number 1 shows the ranking of the most important features, and the higher the ranking number, the importance will decrease until number 9, which has the lowest feature importance to the model. The Bi-LSTM model forecasts have the highest accuracy. The ranking of features from the highest to the lowest is shown as follows: Lag Load, Yearly Hours, Weekdays and Weekend, Compensatory Holiday, Ambient

***Table 6**: Feature Importance Ranking of NN Model by SHAP and MI.*

Feature	Feature importance ranking					
	A	B	C	D	E	F
Lag Load	1	1	1	1	1	1
Yearly Hours	2	2	2	2	2	2
Weekdays and Weekend	3	3	3	3	3	6
Compensatory Holiday	4	4	5	5	5	7
Ambient Temperature	5	5	4	4	4	4
Daytime and Nighttime	9	6	7	8	6	3
Dew Point	6	7	6	6	7	5
Long Weekend	8	8	8	7	8	8
Quarter-Hourly Interval	7	9	9	9	9	9

Temperature, Daytime and Nighttime, Dew Point, Long Weekend, and Quarter-Hourly Interval, respectively, were compared for all 6 models. The Bi-LSTM and LSTM models are models that learn the time sequence of data. The 4th feature is Compensatory Holiday, which shows the compensatory holidays of the university reflect the overall electricity consumption behavior. There are red and blue dots clearly separated from the blue ones. The model can recognize clear patterns with this data and associate it with the load drop. This differs from the fifth-ranked feature, ambient temperature. Since alternating points in the same location have a non-linear relationship, fluctuations can make it more difficult for models to capture these patterns.

A CNN-LSTM model, which captures short-term and long-term patterns well, or an Attention-LSTM model that emphasizes time steps, will better understand this feature. Based on feature priorities, CNN-LSTM and Attention-LSTM models prioritize ambient temperature over compensatory holidays. DNN models can easily learn by finding association weights, do not remember event sequences, and process data without considering time sequences. The SHAP value of the Ambient Temperature feature, used to describe load changes, is higher than that of the Compensatory Holiday feature, which relies on time sequences and previous behaviors. The Quarter-Hourly Interval feature, also a later feature, is considered a fundamental feature, as it does not directly correlate with load and is a simple feature for the model to learn. From the simulation results, it was found that the MI method gives more importance to the Daytime and Nighttime, Ambient Temperature, and Dew Point features of each model than the SHAP method because these features have more raw data statistics related to the forecasting target, which reflects the actual behavior of electrical loads without learning the model mechanism and not understanding the relationship between the features like SHAP. When using SHAP, the importance is reduced. For Weekdays and Weekend and Compensatory Holiday features, which have lower MI scores, SHAP gives higher importance to both features. This is because the model learns the effect of these features through the more important Lag Load feature, not just learning the features alone. Using SHAP, one can learn features with complex relationships better.

The third feature of the MI method, Daytime and Nighttime, is shown to be highly correlated with load. However, compared to the SHAP method, this feature is ranked low due to the model's inability to extract patterns well. However, the key features, Lag Load and Yearly Hours, ranked first and second in both methods, showed consistent results, indicating that this feature set can definitely improve accuracy and efficiency. Using load data, the MI method screens out potentially important features related to electrical loads, screening relevant and related data before importing into the model and isolating truly irrelevant features. The trend in importance of features is evident, eliminating bias from the model for optimal forecasting performance. However, since it is not yet possible to determine how features affect model decisions, SHAP values are used at each data point in real-world situations to explain the relationship between features and model decisions. This ensures that the model is indeed using those features and provides a deeper explanation of model behavior, demonstrating the importance of features that directly influence the model. Furthermore, under different load conditions, the direction of their impact on forecasting can be identified. Regarding the reliability of black-box models, this can be explained.

4. CONCLUSION

This research proposes the development of a deep learning-based forecasting technique. The accuracy of the model is compared with that of a baseline NN model. The Bi-LSTM model is shown to have the highest accuracy and performance for load forecasting. The data collected through cleaning and feature engineering are then transformed into features ready for input into the forecasting model. The electricity load forecasts are most closely aligned with actual values. Overall, the Bi-LSTM and LSTM models also yield the lowest metric values, averaging across all tested data, indicating the model's potential for forecasting in this context.

Furthermore, the model reliability investigation indicates that transparent feature analysis and model description using MI can explain the relationships between raw data and electricity loads and filter static data for inspection before inputting the model. This reduces the number of features required for model input and reduces training time. SHAP provides in-depth model descriptions, demonstrating the actual use of the model's features, understanding complex features, and increasing model reliability because all features have underlying relationships that can be demonstrated. Using this combined approach, comprehensive coverage is achieved. It provides accurate forecasting results and can also provide causal explanations.

ACKNOWLEDGEMENT

This work is fully supported by Suranaree University of Technology.

REFERENCES

[1] International Energy Agency (IEA), World Energy Outlook 2023. Paris, France: IEA, 2023. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2023

[2] U.S. Energy Information Administration, *Electricity explained: Factors affecting electricity consumption*, 2023. [Online]. Available: https://www.eia.gov/energyexplained/electricity/

[3] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.

[4] T. Panapongpakorn and D. Banjerdpongchai, "Short-term load forecast for energy management system using neural networks with mutual information method of input selection," in *Proc. SICE International Symposium on Control Systems (SICE ISCS)*, Tokyo, Japan, 2019, pp. 13–18.

[5] Ministry of Energy Thailand, *Thailand Smart Grid Development Plan (2022–2031)*, Bangkok, Thailand, 2022. [Online]. Available: https://www.energy.go.th

[6] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001.

[7] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLOS ONE*, vol. 13, no. 3, Mar. 2018.

[8] S. Atef and A. B. Eltawil, "Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting," *Electrical Power Systems Research*, vol. 189, Dec. 2020.

[9] Y.-G. Lee, J.-Y. Oh, D. Kim, and G. Kim, "SHAP value-based feature importance analysis for short-term load forecasting," *Applied Science and Convergence Technology*, vol. 18, pp. 579–588, 2023.

[10] J. Jang, W. Jeong, S. Kim, B. Lee, M. Lee, and J. Moon, "RAID: Robust and interpretable daily peak load forecasting via multiple deep neural networks and Shapley values," *Sustainability*, vol. 15, no. 8, 2023.

[11] X. Liu, Z. Yang, Y. Guo, Z. Li, and X. Xu, "A novel correlation features self-assigned Kolmogorov-Arnold Networks for multi-energy load forecasting in integrated energy systems," *Energy Conversion and Management*, vol. 300, 2024.

[12] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," *IEEE Access*, vol. 11, pp. 48322–48331, 2023.

[13] P. Wu, Z. Huang, Y. Pian, L. Xu, J. Li, and K. Chen, "A Combined Deep Learning Method with Attention-Based LSTM Model for Short-Term Traffic Speed Forecasting," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, 2020.

[14] K. Ullah et al., "Short-Term Load Forecasting: A Comprehensive Review and Simulation Study With CNN-LSTM Hybrids Approach," *IEEE Access*, vol. 12, pp. 111858–111881, 2024.

[15] S. Hu et al., "Short-Term Load Forecasting Based on Mutual Information and BI-LSTM Considering Fluctuation in Importance Values of Features," *IEEE Access*, vol. 12, pp. 23653–23665, 2024.

[16] R. Azimi, A. Amjady, and A. Daraeepour, "Mutual Information-Based Inputs Selection for Electric Load Time Series Forecasting," *Entropy*, vol. 15, no. 3, pp. 926–942, Mar. 2013.

[17] D. A. Bolstad, U. Cali, M. Kuzlu, and U. Halden, "Day-ahead Load Forecasting using Explainable Artificial Intelligence," in *Proc. IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, New Orleans, LA, USA, 2022, pp. 1–5.

[18] M. Ramya and T. Devaraju, "Dynamic scheduling and islanding energy management in microgrids using deep learning and adaptive reinforcement learning techniques," in *Proc. IEEE International Conference on Intelligent Systems and Sustainable Innovation (ICISS)*, 2025.

[19] P. Shanmugapriya, D. R. P. R., S. V. S., and M. SureshKumar, "Effective approach for energy management system using machine learning algorithm," in *Proc. IEEE International Conference on Power, Energy, Control and Transportation Systems (ICPECTS)*, 2024.

[20] J. Zhang, S. Shi, and N. Zhang, "Mitigation strategies for greenhouse gas emissions and energy consumption based on deep learning," in *Proc. IEEE International Conference on New Perspectives in Smart Industrial Future (NPSIF)*, 2024.

[21] J. D. Joy, M. R. Hossain and A. Chowdhury, "A Hybrid Deep Learning Model for Day Ahead Short-Term Load Forecasting," 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON), Rajshahi, Bangladesh, 2024, pp. 1-6.

[22] M. A. Hassan, A. A. Eladl, B. E. Sedhom, and M. A. Saeed, "A new optimal centralized demand side management for a campus smart microgrid," in *Proc. IEEE International Symposium on Industrial Electronics (ISIE)*, 2023.

[23] S. Akhtar, M. Z. bin Sujod, and S. S. H. Rizvi, "A novel deep learning architecture for data-driven energy efficiency management (D2EEM) – systematic survey," in *Proc. IEEE International Conference on Engineering and Emerging Technologies (ICEET)*, 2021.

Nisachon Thabchareceived her Bachelor of Engineering degree in Electrical Engineering (Second Class Honors) from Suranaree University of Technology, Thailand in 2024. She is currently pursuing her Master's degree in Electrical Engineering at the Faculty of Electrical Engineering, Institute of Engineering, Suranaree University of Technology. Her current research interests focus on power systems and the application of artificial intelligence in power systems.

Nitikorn Junhuathon was born in Nakhon Ratchasima in Thailand, on February 23, 1995. He received his B. Eng., M. Eng., and Ph.D. In Electrical Engineering from Suranaree University of Technology (SUT), Thailand, in 2017, 2019 and 2023, respectively. He is currently a lecturer at Department of Electrical Engineering, Faculty of Engineering, Raja Mangala University of Technology Thanyaburi (RMUTT), Thailand. His research interests are machine learning and optimization technique for power forecasting, electric vehicle application, and related fields.

Keerati Chayakulkheeree received his B. Eng. in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL) in 1995, and M. Eng. and D. Eng. degrees in Electric Power System Management from the Asian Institute of Technology (AIT), Thailand, in 1999 and 2004, respectively. He is an Associate Professor at the School of Electrical Engineering, Institute of Engineering, Suranaree University of Technology (SUT), Nakhon Ratchasima, Thailand. His research interests include power system optimization, AI applications in power systems, smart grid technology, power system protection, distributed energy resources, electricity market restructuring, renewable energy integration, and carbon emissions in the power sector.