INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 6 (2025) 58-70
https://publisher.uthm.edu.my/ojs/index.php/ijie

Application of Artificial Intelligence in Short-Term Load Forecasting at Low-Voltage Substations

Quoc Van Nguyen[1,2], Thang Viet Tran[3], Ngo Minh Tri Nguyen[3], Chi-Ngon Nguyen[1]*

[1] *College of Engineering,*
Can Tho University, Can Tho City, 900000, VIETNAM
[2] *HUTECH Institute of Engineering,*
HUTECH University, Ho Chi Minh City, 700000, VIETNAM
[3] *Institute of Semiconductor Microchip Technology,*
Nguyen Tat Thanh University, Ho Chi Minh City, 700000, VIETNAM

*Corresponding Author: ncngon@ctu.edu.vn
DOI: https://doi.org/10.30880/ijie.2025.17.09.05

Article Info

Received: 27 June 2025
Accepted: 12 December 2025
Available online: 31 December 2025

Keywords

Short-term load forecasting, artificial intelligence, deep neural network, artificial neural network

Abstract

This study presents a Short-Term Load Forecasting (STLF) process in Vietnam. A total of 13 features, including 9 electricity-related features and 4 time-related features, are extracted to predict the Total Active Power (TAP) of a three-phase low-voltage transformer station. The Mutual Information (MI) analysis results show that the Active Power (AP) features in the three phases have a significantly higher importance level than the other features in predicting TAP. Two algorithms, including Deep Neural Network (DNN) and Artificial Neural Network (ANN), are used to predict TAP. In addition to the two models using all 13 features, two corresponding models using only 3 AP features are also trained for comparison. Results show that, with 17 047 data points, models using 3 AP features have a suitable level of complexity for better results than similar models using all 13 features. The DNN model with 3 features yields the best results, with MAE, RMSE, and R2 values of 1.9089, 4.2880, and 0.9971, respectively. In future research, the number of data points will be improved to explore better features, and other features could also be considered.

1. Introduction

In Vietnam, the electrical system is divided into three main voltage levels, reflecting the function and scope of each level. High voltage (above 35 kV) is primarily employed in the transmission grid, exemplified by the 220 kV and 500 kV North-South lines, playing a key role in transmitting large-scale power from generation plants to load centers and regional substations. Medium voltage (above 1 kV up to 35 kV), with 22 kV being the most commonly used level, is applied in the distribution network to supply electricity to urban areas, industrial zones, as well as medium and large loads, through 110/22 kV substations. Low voltage (not exceeding 1 kV), typically 0.4 kV, is mainly used to deliver electricity directly to households, commercial buildings, and small-scale production facilities. The increasing presence of small-scale loads and distributed energy resources is reshaping low-voltage power grids. As a result, load forecasting has become indispensable, supporting both operational activities and the overall efficient management of the energy system [1, 2].

Load forecasting plays a crucial role in enabling electric utilities to plan their finances, manage electricity market prices, and efficiently plan generation, supply, and dispatch based on load fluctuations. This not only

ensures the safety, security, and stability of the smart grid system but also optimizes energy management and reduces overall operating costs, thereby improving the efficiency and sustainability of the energy sector [3-6].

Distribution load forecasting is typically categorized into short-term, medium-term, and long-term projections. Short-Term Load Forecasting (STLF), specifically, concentrates on predicting load demand over periods ranging from several hours to a week. This capability is indispensable for ensuring the reliable and efficient operation of increasingly prevalent smart grid systems by optimizing power distribution and supply [7]. STLF provides crucial information for decision-making in meeting customer electricity demand, reducing the risk of power outages, and optimizing the response to electricity demand. Additionally, it supports load demand monitoring, peak load reduction, power outage detection and restoration, as well as intelligent customer information management. Therefore, achieving high accuracy in STLF is a necessary challenge for electricity businesses and suppliers to effectively coordinate when participating in the electricity market [8, 9]. Many traditional load forecasting methods such as time series analysis, regression analysis, extrapolation, and elasticity coefficient have been applied, however, they often do not achieve high accuracy [10-14].

In recent years, computer science and Artificial Intelligence (AI) have experienced rapid development and have been applied in numerous fields, including electrical systems [15-17]. In particular, they have been utilized in electric load forecasting with the expectation of improving accuracy. Recent research has demonstrated that AI-powered approaches outperform traditional methods in forecasting electricity load due to their superior capacity to process intricate, nonlinear data with greater adaptability [18-22].

Machine learning algorithms such as ANN, DNN, and Fuzzy Neural Network (FNN), have been applied to electric load forecasting and have demonstrated high performance and accuracy [23-26]. Support Vector Machines (SVM), Random Forest Regression (RFR), and Long Short-Term Memory (LSTM) were used for short-term load forecasting. An ensemble method was proposed to improve accuracy by combining the advantages of these models [27]. DNN, Feed Forward Deep Neural Network (FF-DNN), and Recurrent Deep Neural Network (R-DNN) have been employed for short-term load forecasting [28]. FF-DNN and R-DNN have been utilized in conjunction with Time-Frequency (TF) feature selection to enhance forecasting performance [29].

The objectives of this research are (1) to analyse real-time load data collected from the central system, monitor, supervise, and manage real-time data at the Phu My station of Tan Thuan Power Company in Ho Chi Minh City, Vietnam; and (2) to develop DNN and ANN models for STLF. These objectives also enable the identification of suitable features and the selection of appropriate forecasting models, ensuring the robustness and accuracy of the prediction process. The outcomes serve to optimize grid operation, enhance the reliability of power supply, support decision-making in energy resource management, and effectively coordinate power supply for each region.

2. Materials and Methods

2.1 Data Collection

Data was collected at the Phu My substation, rated at 560 kVA - 22/0.4 kV, within the low-voltage distribution network of Tan Thuan Power company, Ho Chi Minh City, Vietnam in real-time from the monitoring and data management system. Data was collected every 30 minutes, starting from 1:30:00 on January 14th, 2024, and ending at 04:30:00 on January 3rd, 2025. After being collected, the data is preprocessed, then used for load forecasting. The entire processing diagram is illustrated in Fig. 1. The processing is conducted in Python.

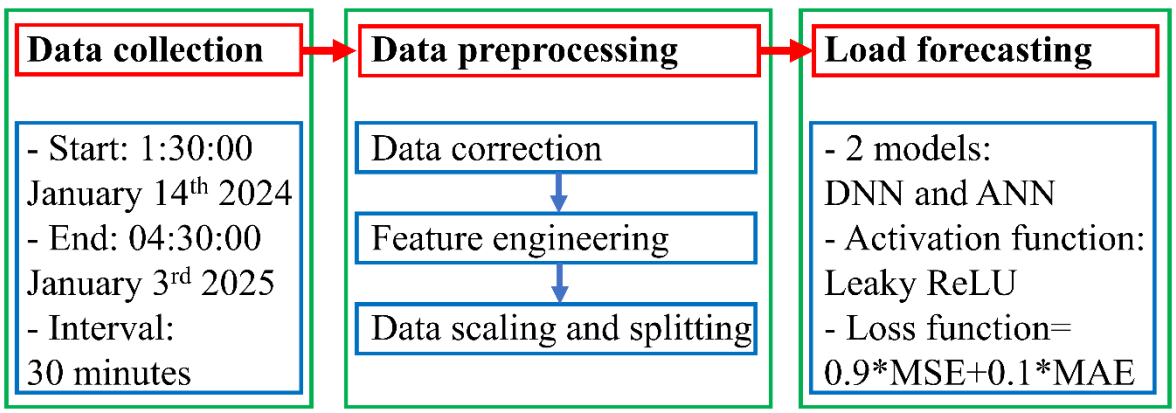

Fig. 1 *Processing diagram*

Int. Journal of Integrated Engineering Vol. 17 No. 9 (2025) p. 58-70

2.2 Data Preprocessing

2.2.1 Data Correction

After being collected, data undergoes preprocessing. The operation and data collection process of the power grid can sometimes encounter incidents, causing some data points to be missing, wrong, duplicated, or lacking certain features. Preprocessing is firstly aimed at addressing these issues, including handling unsuitable values, removing duplicated data, filling missing data points in time intervals.

Handling unsuitable values: Not a Number (NaN) value is unsuitable. Negative or 0 values in case of active power (AP) are also unsuitable. 0 value theoretically can appear in case of reactive power (RP), but is hardly, so is also considered as unsuitable. These unsuitable values should be replaced. Unsuitable values of AP are replaced by mean of positive active values within the same hour. If there is no positive value of this hour, the mean of all positive values is used instead. Meanwhile, unsuitable values of RP are recalculated from apparent power and power factor. If one of these two values are not available, the unsuitable values of RP are tackled similar to that of active values. Please note that apparent power will not be extracted as features in later steps.

Removing duplicated data: When the system experiences a sudden power outage followed by immediate restoration (within a few seconds, for example), an additional data point is recorded that is similar to the previously recorded data point. In this process, if duplicate data points are detected, the first data point is kept. If there are missing feature in this data point, they are filled by the respective features in other data point(s) if available.

Filling missing data points in time intervals: In this study, data are collected each 30 minutes. However, due to system operation problems, data in the times series can be missed. The missing data are filled based on gap size. For gaps within 2 hours, linear interpolation is used. For gaps higher than 2 hours but not higher than 6 hours, cubic interpolation used instead. For gaps higher than 6 hours, data is filled by that of the previous week.

2.2.2 Feature Engineering

The data in this study were collected from the actual operation of the power company. The system continuously records electrical values over time. In this study, both electrical-related and time-related features are employed to diagnose the Total Active Power (TAP) of the three-phase low-voltage transformer station.

Electrical-related features used in this study include AP, RP, and the phase angle of each phase in the three-phase system. These are fundamental quantities that directly reflect the process of power transmission and consumption. AP represents the portion of energy actually utilized to perform work, RP reflects the energy exchange between source and load caused by inductive or capacitive characteristics of equipment, while the phase angle describes the displacement between voltage and current, thereby indicating the efficiency of operation and the degree of phase imbalance in the system. The simultaneous combination of these 3 features across all 3 phases enables the model to capture the complete operating state of the low-voltage transformer, from the magnitude of energy consumption to the quality of power. In total, 9 electrical-related features are extracted.

The primary data was collected in 2024 (starting from January 14, 2024), with a small portion collected in early 2025 (until January 3, 2025). There is no overlap in the dates between the two years. Additionally, there is no significant difference between the data from the 2 years; therefore, the Year feature was excluded from this study. Information regarding the Month was retained as it is relevant to electricity usage patterns, such as weather and seasonal activities.

Information about the day of the month has no significant correlation with electricity usage and is therefore eliminated. Conversely, information about the day of the week is extracted as it is related to industrial, economic, and other activities. In Python, days from Monday to Sunday are encoded from 0 to 6. Since the day of the week is cyclical, and to avoid algorithms mistakenly considering Monday and Sunday as being far apart, the information about the day of the week is encoded into 2 features: “Day of week sine” and “Day of week cosine”. Similarly, information about Hour, Minute, and Second is extracted into 2 features: “Hour sine” and “Hour cosine”. The time encoding using sine and cosine is expressed by the following formulas:

$$t_{sin} = sin\left(\frac{2\pi t}{T}\right) \tag{1}$$

$$t_{cos} = cos\left(\frac{2\pi t}{T}\right) \tag{2}$$

Where, t denotes the time variable, specifically either Day of week or Hour. T represents the corresponding period of t, more specially., 7 days for Day of week and 24 hours for Hour. t_{sin} and t_{cos} are the sine and cosine encoded values of t, respectively. The plots of these encoded features are illustrated in Fig. 2. A value of 7 for Day of week corresponds to Monday of the following week. Similarly, a value of 24 for Hour corresponds to 0:00 of the

next day. In total, 4 time-related features are extracted. Together with the 9 electrical-related features, a total of 13 features are used to predict TAP.

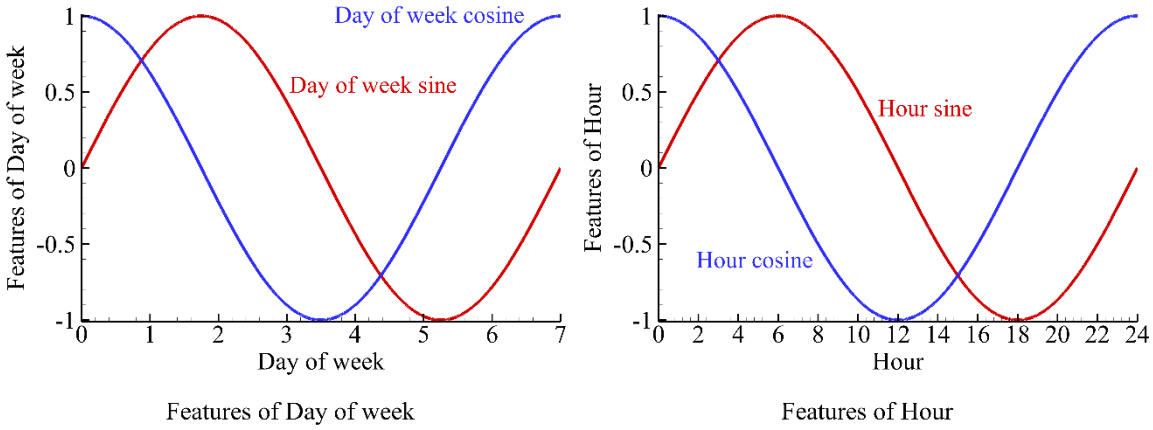

Fig. 2 *Encoded time-related features*

2.2.3 Data Scaling and Splitting

Following these preprocessing steps, the data is scaled to similar ranges. Since the features employed in this study have significantly varying scales, feature scaling accelerates the convergence process during gradient descent-based model training. Standardization is chosen as the feature scaling method. The data is divided into training (70%) and test (30%) sets.

2.2.4 Feature Importance

To identify the most valuable features influencing the prediction, the mutual information (MI) method was adopted for feature importance analysis. MI quantifies the amount of shared information between each feature and the label, thereby measuring their statistical dependency. Unlike linear correlation coefficients, MI is capable of capturing nonlinear relationships between variables, which are common in real-life problems. By computing MI scores for all features, the analysis highlights those with the strongest dependency on the label. In this study, MI was estimated using the KNN-based estimator proposed by Kraskov [30].

2.3 Load Forecasting

2.3.1 Model Structure

In this study, DNN and ANN were used for load forecasting. DNNs and ANNs are commonly used for load forecasting, demonstrating good performance [23-26, 31], and are used in this study. DNN can be considered as a more complex version of ANN with more hidden layers and has been proven to be more effective than traditional methods [32, 33].

In this study, the DNN model consists of 6 layers: 1 input layer, 3 hidden layers, and 1 output layer. The input layer has 13 nodes, corresponding to 13 features, and the output layer has 1 node for predicting load power. The 4 hidden layers have 64, 32, and 16 nodes respectively (from input to output). The ANN model is similar, but it only has 2 hidden layers (the first 2 hidden layers of the DNN model). The ANN model was chosen to be not significantly different from the DNN model to facilitate the assessment of the relationship between model complexity and performance.

In neural-network-based architectures, the number of hidden layer nodes is often chosen as powers of two (for example, 16, 32, 64, 128, 256). Although this is not a strict requirement, it has become a common practice for two main reasons. Firstly, parallel processors such as GPUs and TPUs are optimized to operate on data blocks whose sizes are divisible by two, thereby improving computational efficiency and reducing memory overhead. Secondly, many representative network architectures are designed with a doubling or halving rule for the number of nodes to maintain consistency between feature compression and expansion [34]. For instance, VGGNet [35] implements feature channels of 64–128–256–512, ResNet [36] adopts 64–128–256–512, while U-Net [37] expands up to 1024 channels before symmetrically reducing them. Therefore, selecting hidden nodes as powers of two has become a practical convention in modern deep learning model design.

In ANN and DNN models applied to prediction tasks, the number of hidden nodes typically decreases across layers, or in some cases remains constant at certain intermediate layers before eventually decreasing. This design

reflects the principle of feature compression, in which the initial layers extract and represent high-dimensional input data, while subsequent layers focus on more relevant information and discard noise. Gradually reducing the number of nodes not only improves generalization and mitigates overfitting, but also lowers computational cost by reducing the number of trainable parameters. Consequently, the principle of gradual reduction, or maintaining and then reducing nodes, has become a common design practice in ANN and DNN prediction models.

Based on experimental results with our dataset, the models with two or four hidden layers, corresponding to the ANN and DNN in this study, achieved the best performance. The number of nodes was also tested with different settings, but the outcomes were either worse or not significantly better; therefore, the standard configuration was adopted.

Rectified Linear Units (ReLU) are commonly used as activation functions due to their fast convergence. However, if the input is always negative, it can lead to the "dying ReLU" problem, where the corresponding node ceases to learn. A variant of ReLU, Leaky ReLU, uses a small slope for the negative part of the input to address this issue. In this study, Leaky ReLU is used as the activation function in the hidden nodes.

2.3.2 Loss Function

Mean Squared Error (MSE) and Mean Absolute Error (MAE) are often used to evaluate the difference between predicted values and actual values of data. MSE and MAE are calculated based on the following formulas:

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2 \tag{3}$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|\left(y_j - \hat{y}_j\right)\right| \tag{4}$$

Where, n signifies the sample size, y_j is the actual outcome for the j-th data point, and $\hat{y}_j$ is the model's prediction for the same data point.

Both of these metrics compare predicted values to actual values of the data. However, while MAE calculates the absolute value of the difference, MSE calculates the square of the difference between the two values. This causes MSE to penalize larger differences more than MAE, but also makes MSE more sensitive to outliers [38].

Overall, MSE offers better overall benefits, but MAE should also be considered to increase the model's stability. Therefore, several hybrid approaches combine these two metrics to leverage the advantages of both [38-41]. In this study, the DNN's loss function is a sum of 90% MSE and 10% MAE. This combination helps the model perform well under normal conditions, but is more stable when anomalies occur.

2.3.3 Evaluation Metrics

A total of three metrics was used to evaluate the models, including MAE, Root Mean Square Error (RMSE) and R-squared (R^2). These metrics are specifically designed to measure the accuracy of continuous variables and are commonly applied in the field of electricity load forecasting. The loss function is not used for evaluation because it is difficult to assess and also difficult to compare results with other models.

MAE has been mentioned in equation (3), while RMSE is simply the square root of MSE in equation (4). RMSE is used instead of MSE because it has the same unit as MAE and the predicted quantity (kW). R^2 is calculated based on the following formula:

$$R^2 = 1 - \frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{\sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2} \tag{5}$$

Where, n signifies the sample size. y_j is the actual outcome for the j-th data point, $\hat{y}_j$ is the model's prediction for the same data point, and $\bar{y}$ is the mean of y_j. R^2 represents the proportion of variance that the predicted model explains.

3. Results and Discussion

3.1 Active power distribution

Fig. 3 shows the Active Power (AP) distribution of the system at 3 phases after data correction, comprising a total of 17,407 data points (time is labeled in the format “yyyy-mm,” e.g., “2025-01” denotes January 2025). Analyzing

the processed power profiles is a crucial step in assessing data quality and identifying any remaining anomalies or patterns of interest. Please note that the corresponding stick of a month is placed at the beginning of that month. Thus, the time period of each month in the figure is shown from the stick of that month to the stick of the following month.

From early 2024 until early February, the power capacity was relatively stable, fluctuating between 50-125 kW. From March to May 2024, the capacity increased significantly, at times exceeding 175 kW. This correlation could be attributed to the peak temperatures in Ho Chi Minh City occurring around April, leading to a surge in demand for air conditioners and fans. From June to August 2024, the capacity decreased and stabilized, ranging from 25-100 kW. In the last quarter of 2024, the capacity increased again and fluctuated significantly between 50-150 kW.

The AP profiles of phases A, B, and C are depicted in red, green, and blue on the graph. Phase C demonstrates the highest power magnitude and significant variability, whereas phase B exhibits the lowest power level. Phase A maintains the most consistent power output. A notable phase imbalance is observed during the peak months of July, August, and September. The occurrence of zero power points may be attributed to system faults or scheduled maintenance.

The data in this study was collected from 01:30:00 on January 14th, 2024, to 04:30:00 on January 3rd, 2025, covering a period of less than one year. This limits the ability to analyze and predict TAP in relation to annual seasonal patterns at the intervals where data is missing. This shortcoming should be addressed in future work by improving the data collection process.

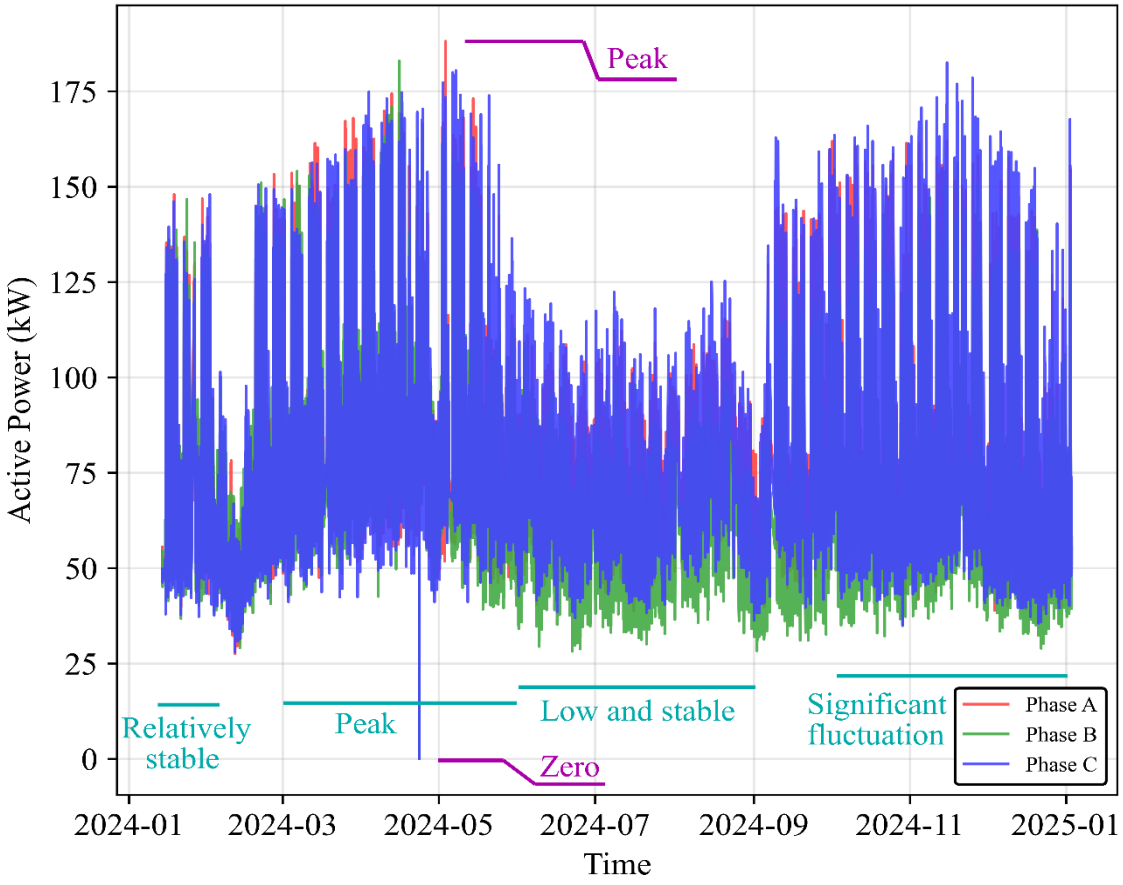

Fig. 3 *Active power distribution at 3 phases*

3.2 Feature Importance

Fig. 4 demonstrates the relative importance of 13 features based on MI. AP of the 3 phases exhibits the highest significance compared to other features, indicating that they are the most crucial predictors of load.

The next most important features are reactive power, angle in the three phases, and features related to the hour. Reactive Power (RP) and Angle have similar values in representing AP in the same phase, thus their importance levels are similar. The electrical features related to phase C have a higher level of importance compared to the other two phases.

Int. Journal of Integrated Engineering Vol. 17 No. 9 (2025) p. 58-70

Hour-related features are important because they are related to daily electricity usage patterns. Meanwhile, features related to the day of the week have the lowest level of importance. This result suggests that the level of electricity consumption is less related to the day of the week, but has a significant dependence on the time of day.

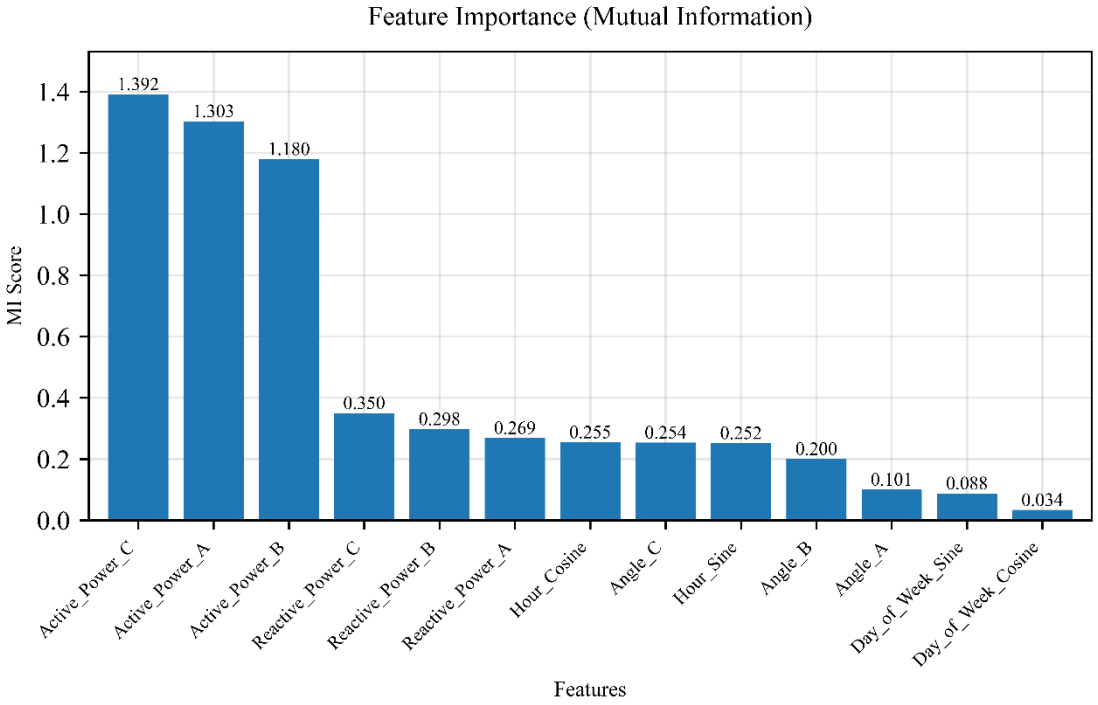

Fig. 4 *Feature importance (mutual information) of 13 features*

Fig. 5 displays a correlation matrix of 14 variables, including 13 features and TAP, visualized as a heatmap for better clarity. TAP has a very strong correlation with AP-C (0.96), followed by AP-A (0.96) and AP-B (0.95). This result aligns with Fig. 4, where the 3 AP values are shown to be crucial for predicting TAP.

In an ideal three-phase system, the AP values in each phase are balanced to optimize system performance. However, in practice, due to load variations, these values often exhibit small discrepancies. In this study, the AP values also showed a high correlation with each other (0.88-0.90).

This study uses data from low-voltage substations to ensure practical relevance, rather than building a dedicated data acquisition system. However, these substation datasets only record electrical information and lack temperature data. In Ho Chi Minh City, Vietnam, the annual temperature varies within a narrow range of approximately 28 °C–33 °C, yet it can still affect electricity demand for devices such as air conditioners and fans. In the future, a temperature monitoring system will be developed and integrated into the existing data collection framework, allowing both its utilization and the evaluation of its potential as a feature for TAP prediction.

3.3 Load Patterns

Fig. 6 illustrates the daily and weekly load patterns of the system, spanning from 00:00 to 23:00 hours and from Monday to Sunday, respectively. From 00:00 to 06:00, corresponding to people's sleep time, the power consumption is at its lowest, followed by a rapid increase. During working hours from 08:00 to 17:00, the power consumption is high, with two peaks at 11:00 and 14:00. The period from 12:00 to 13:00 represents the trough between the two peaks, corresponding to lunchtime. After 16:00, the power consumption starts to decrease rapidly and returns to the initial low state. It can be seen that from Monday to Friday, the levels of electricity consumption are similar, then decrease on Saturday and continue to decline on Sunday. This is related to the working culture in Vietnam, where most employees work from Monday to Friday, with a considerable proportion working either full day or only in the morning on Saturday, and the majority having Sunday off. The above charts show a significant relationship between Hour and Day of the week with electricity consumption levels. Therefore, the features related to them may contain useful information for predicting TAP.

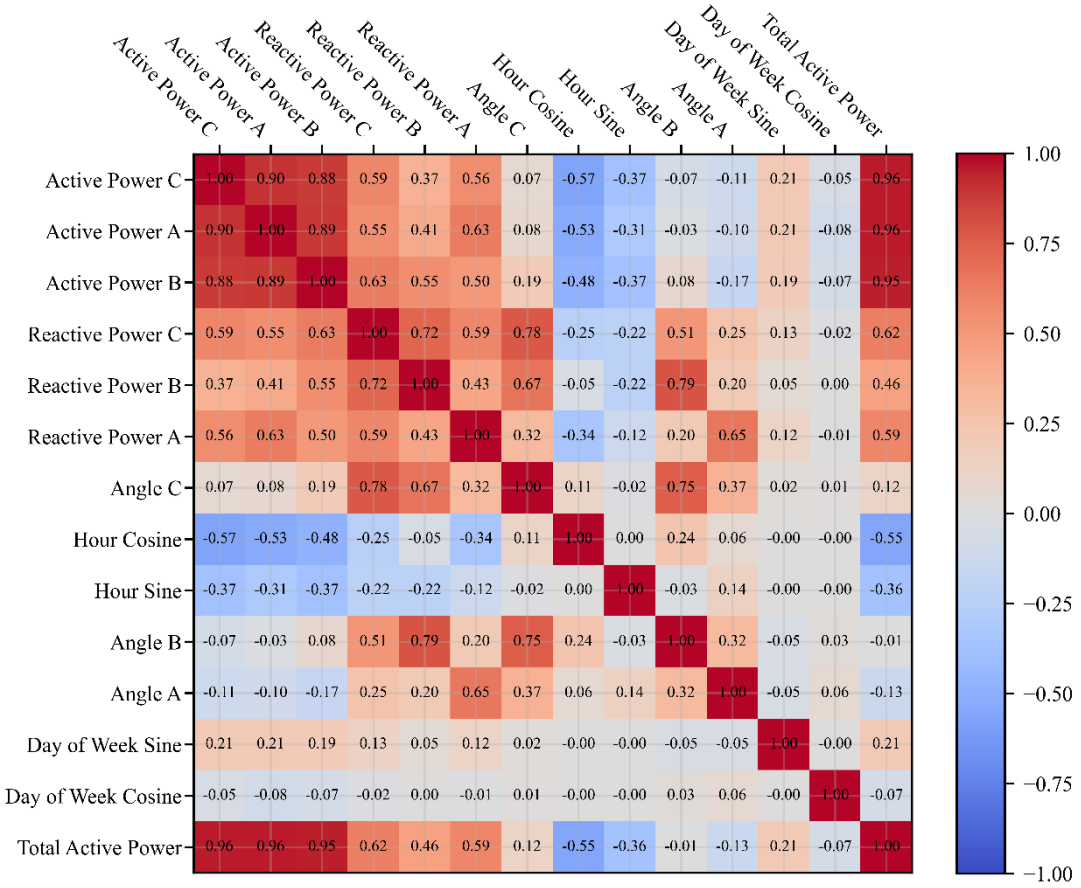

Fig. 5 *Correlation matrix of 13 features and total active power*

3.4 Model Evaluation

Besides the aforementioned DNN and ANN models, 2 other similar models were trained and tested for comparison. These models have similar structures to the original two, but only have 3 nodes in the input layer, corresponding to the 3 features with the highest MI indices, which are AP in the three phases. In total, four models were developed, representing all combinations of the two algorithm types (DNN and ANN) and the two feature groups: one with 13 features (AP, RP, and Angle of each phase; Hour sine, Hour cosine, Day of week sine, and Day of week cosine), and one with 3 features (AP of the three phases).

Table 1 presents the performance of the 4 TAP prediction models. The R^2 values of all models are not significantly different and are no less than 0.9965. This indicates that all models can predict the load relatively effectively.

DNN model with 3 features achieved lower MAE and RMSE compared to the DNN with 13 features. This can be explained by the fact that DNN has a complex structure with many parameters, which makes it more prone to overfitting when the number of features is large while the training data is limited. Reducing the number of features makes the model more compact, enhances generalization ability, and consequently improves prediction accuracy.

For ANN, due to its simpler structure and fewer parameters, it is less sensitive to overfitting compared to DNN. With 13 features, ANN can take advantage of the additional information to reduce RMSE, although its MAE remains higher. Conversely, with 3 features, ANN maintains a lower MAE but results in a higher RMSE because the model cannot fully capture the characteristics of the data.

Comparing the 2 models, DNN with 3 features produces the best results. This is because DNN, with its deeper architecture and larger number of parameters, has a stronger capacity to learn and extract complex features from the data. When the number of features is reduced, DNN is able to effectively utilize the essential information while avoiding overfitting under limited data conditions, thus achieving the lowest MAE and a competitive RMSE. In contrast, ANN, being simpler and with fewer parameters, has less flexible data representation ability; increasing the number of features helps ANN improve RMSE but it cannot achieve the same consistent accuracy as DNN. Therefore, DNN demonstrates superior performance in the limited-data context of this study.

Int. Journal of Integrated Engineering Vol. 17 No. 9 (2025) p. 58-70

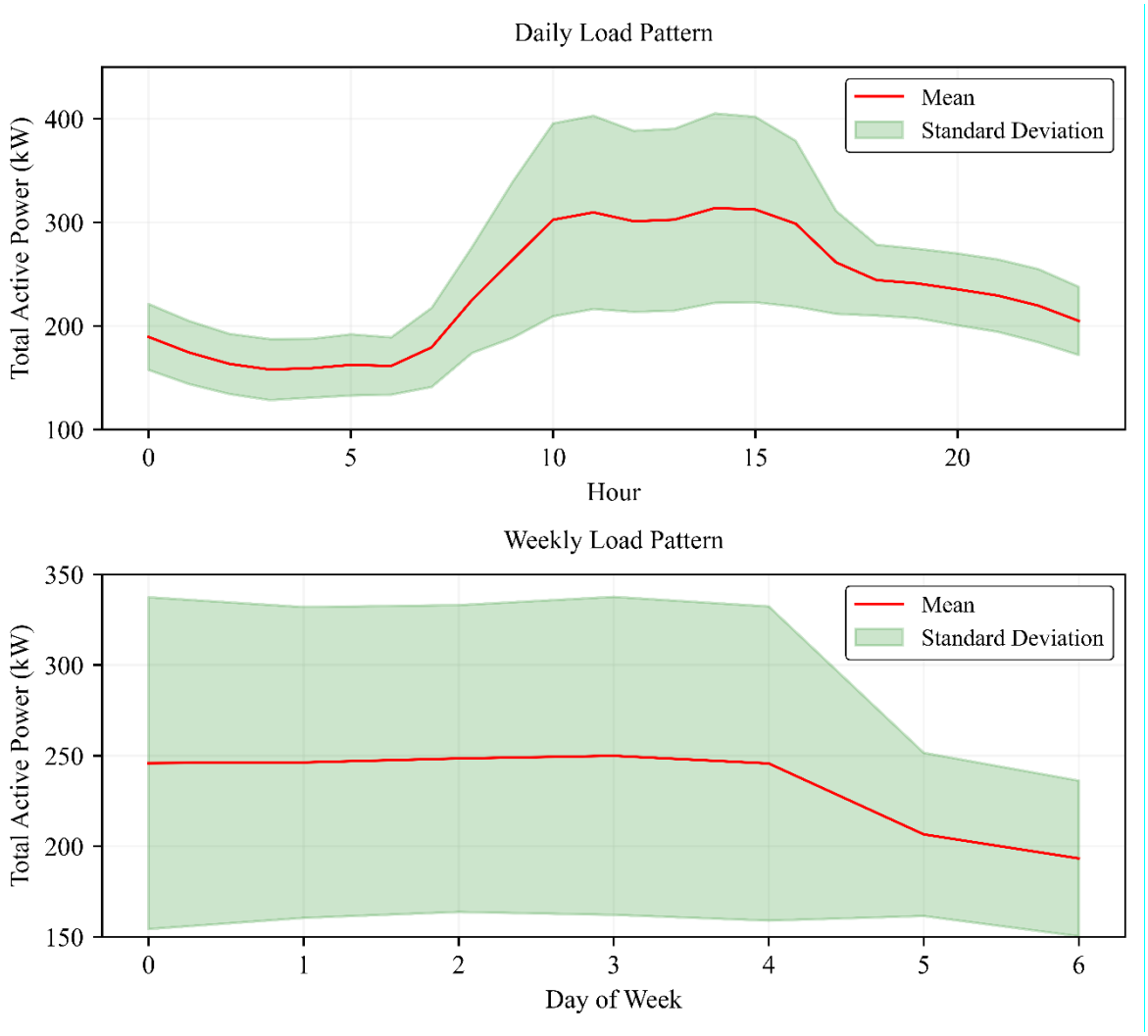

Fig. 6 *Daily and weekly load patterns*

Overall, the DNN with 3 features can be regarded as the best model, since it achieves the lowest MAE (1.9089), a competitive RMSE (4.2880), a similar R2 (0.9971), and avoids overfitting thanks to the reduced number of features. Meanwhile, ANN with 13 features achieves a lower RMSE but at the cost of a higher MAE, indicating less stable performance.

For reference, several studies have reported the following results. A hybrid ensemble deep learning (HEDL) model for load forecasting achieved MAE = 4.24 and RMSE = 6.25 under the best conditions [42]. A sensitivity-enhanced recurrent neural network (SERNN) combined with particle swarm pptimization (PSO) and Bayesian optimization (BO) for hosting capacity forecasting reported RMSE values of 0.38 and 0.31, with R^2 scores of 0.97 and 0.98 on the IEEE-123 network and a real Australian dataset [40]. Using household load data, the clustering-based LSTM model achieved the best performance, with an MAE of 0.0033 and an RMSE of 0.054, owing to the repetitive and predictable nature of household consumption patterns [39]. However, these metric values should be considered only as reference points, since they strongly depend on the chosen approach and the diversity of the datasets.

Table 1 Performance of TAP prediction models

Metric	DNN models		ANN models		Metric note
	13 features	3 features	13 features	3 features	
MAE	2.2968	1.9089	2.5905	2.1183	Lower is better
RMSE	4.4035	4.2880	4.1777	4.6091	Lower is better
R^2	0.9969	0.9971	0.9965	0.9966	Higher is better

4. Conclusions

The study collected real-world data from a low-voltage substation operating in Vietnam and applied multiple preprocessing steps to address data issues arising from actual operational conditions, rather than relying on pre-cleaned or publicly available datasets. Furthermore, incorporating 10% MAE (alongside 90% MSE) in the loss function helps mitigate the model's sensitivity to outliers under real-world conditions. This approach emphasizes practical engineering problem-solving rather than purely academic experimentation.

A total of 13 features were selected for TAP prediction, including per-phase AP, RP, and Angle (9 features); Hour sine, Hour cosine, Day-of-week sine, and Day-of-week cosine. Due to the large number of features and their highly nonlinear relationships, feature importance was evaluated using Kraskov's method instead of simple correlation analysis.

Ideally, three-phase loads should be balanced to ensure efficient operation. However, there are often disparities between phases due to load variations. The AP distribution analysis indicates that phase A has the highest load contribution, while phase B has the lowest. Feature importance analysis based on MI also shows that electrical features in phase C are more significant than those in the other phases. Nevertheless, the pairwise correlation between phases remains high, ranging from 0.88 to 0.90.

MI analysis reveals that 3 AP features have significantly higher predictive value for TAP compared to the remaining features. These 3 features also exhibit a very high correlation with TAP (0.95-0.96) and with each other pairwise (0.88-0.90). Model evaluation results also show no significant difference in R^2 values between models using 13 and 3 features. Given 17047 data points and the corresponding model complexity, the DNN model with 3 features achieves the best performance, with MAE, RMSE, and R^2 values of 1.9089, 4.2880, and 0.9971, respectively. For comparsion, R^2 of related sudies, are 0.918 with generalized linear model-based load forecasting benmark model [43], 0.984 with hybrid deep learning framework [44], and 0.965 with multivariate polynomial regression model [45]. Therefore, with the current number of data points, features other than AP contribute more to noise than to load forecasting. In future studies, the number of data points needs to be improved to better evaluate the features. Additionally, other features could be considered for use.

Acknowledgement

Data of this research is collected with the support of Tan Thuan Power Company, Ho Chi Minh City Power Corporation (EVNHCMC).

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Quoc Van Nguyen, Chi-Ngon Nguyen;* ***data collection:*** *Quoc Van Nguyen, Thang Viet Tran;* ***analysis and interpretation of results:*** *Quoc Van Nguyen, Ngo Minh Tri Nguyen;* ***draft manuscript preparation:*** *Quoc Van Nguyen, Chi-Ngon Nguyen, Thang Viet Tran, Ngo Minh Tri Nguyen. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Gilbert Ciaran, Jethro Browell, & Bruce Stephen (2023) Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks. *Sustainable Energy, Grids and Networks. 34*, 100998 https://doi.org/10.1016/j.segan.2023.100998

[2] Pati Uttamarani, Papia Ray, & Arvind R. Singh (2022) An intelligent approach towards very short-term load forecasting. *International Journal of Emerging Electric Power Systems. 23(1)*, 59-72 https://doi.org/10.1515/ijeeps-2021-0012

[3] Raza Muhammad Qamar, Mithulananthan Nadarajah, Duong Quoc Hung, & Zuhairi Baharudin (2017) An intelligent hybrid short-term load forecasting model for smart power grids. *Sustainable Cities and Society. 31*, 264-275. https://doi.org/10.1016/j.scs.2016.12.006

[4] Sun X., et al. (2016) An Efficient Approach to Short-Term Load Forecasting at the Distribution Level. *IEEE Transactions on Power Systems. 31(4)*, 2526-2537. https://doi.org/10.1109/TPWRS.2015.2489679

[5] Panda Saroj & Papia Ray (2021) Analysis and evaluation of two short-term load forecasting techniques. *International Journal of Emerging Electric Power Systems. 23.* https://doi.org/10.1515/ijeeps-2021-0051

[6] Manz Devon, et al. (2014) The grid of the future: Ten trends that will shape the grid over the next decade. *IEEE Power and Energy Magazine. 12(3)*, 26-36. https://doi.org/10.1109/MPE.2014.2301516

[7] Hong Ye, et al. (2020) A deep learning method for short-term residential load forecasting in smart grid. *IEEE Access. 8*, 55785-55797. https://doi.org/10.1109/ACCESS.2020.2981817

[8] Erişen Esra, Cem Iyigun, & Fehmi Tanrısever (2017) Short-term electricity load forecasting with special days: an analysis on parametric and non-parametric methods. *Annals of Operations Research*, 1-34 https://doi.org/10.1007/s10479-017-2726-6

[9] Xinyu Wu, Dou Chunxia, & Yue Dong (2021) Electricity load forecast considering search engine indices [J]. *Electric Power Systems Research. 199.* https://doi.org/10.1016/j.epsr.2021.107398

[10] Taylor James W & Patrick E McSharry (2007) Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems. 22(4)*, 2213-2219 https://doi.org/10.1109/TPWRS.2007.907583

[11] Nguyen Hung & Christian K Hansen (2017) Short-term electricity load forecasting with Time Series Analysis. *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 214-221 https://doi.org/10.1109/ICPHM.2017.7998331

[12] Zhang Ning, Zhiying Li, Xun Zou, & Steven M Quiring (2019) Comparison of three short-term load forecast models in Southern California. *Energy. 189*, 116358. https://doi.org/10.1016/j.energy.2019.116358

[13] Shahare Kamini, et al. (2023) Performance analysis and comparison of various techniques for short-term load forecasting. *Energy Reports. 9*, 799-808. https://doi.org/10.1016/j.egyr.2022.11.086

[14] Ceperic Ervin, Vladimir Ceperic, & Adrijan Baric (2013) A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems. 28(4)*, 4356-4364 https://doi.org/10.1109/TPWRS.2013.2269803

[15] Hafdaoui Hichem, Salim Bouchakour, & Nasreddine Belhaouas (2022) Using machine learning for analysis a database outdoor monitoring of photovoltaic system. *International Journal of Integrated Engineering. 14(6)*, 275-280. https://doi.org/10.30880/ijie.2022.14.06.024

[16] Setiawan Andi, Bayu Rudiyanto, Satryo Budi Utomo, & Muji Setiyo (2021) Characteristic of Fuzzy, ANN, and ANFIS for Brushless DC Motor Controller: An Evaluation by Dynamic Test. *International Journal of Integrated Engineering. 13(6)*, 274-284. https://doi.org/10.30880/ijie.2021.13.06.024

[17] Uddin Syed Mohammed & Mohd Izhar A Bakar (2023) Intelligent voltage sag compensation using an artificial neural network (ANN)-Based dynamic voltage restorer in MATLAB Simulink. *International Journal of Integrated Engineering. 15(7)*, 249-261. https://doi.org/10.30880/ijie.2023.15.07.023

[18] Bendaoud Nadjib Mohamed Mehdi & Nadir Farah (2020) Using deep learning for short-term load forecasting. *Neural computing and applications. 32(18)*, 15029-15041. https://doi.org/10.1007/s00521-020-04856-0

[19] Kwon Bo-Sung, Rae-Jun Park, & Kyung-Bin Song (2020) Short-term load forecasting based on deep neural networks using LSTM layer. *Journal of Electrical Engineering & Technology. 15*, 1501-1509. https://doi.org/10.1007/s42835-020-00424-7

[20] Shi Zhongtuo, et al. (2020) Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions. *Applied Energy. 278*, 115733. https://doi.org/10.1016/j.apenergy.2020.115733

[21] Rafi Shafiul Hasan, Shohana Rahman Deeba, & Eklas Hossain (2021) A short-term load forecasting method using integrated CNN and LSTM network. *IEEE access. 9*, 32436-32448. https://doi.org/10.1109/ACCESS.2021.3060654

[22] Ullah Fath U Min, et al. (2022) Deep Learning-Assisted Short-Term Power Load Forecasting Using Deep Convolutional LSTM and Stacked GRU. *Complexity. 2022(1)*, 2993184. https://doi.org/10.1155/2022/2993184

[23] Tzafestas Spyros & Elpida Tzafestas (2001) Computational intelligence techniques for short-term electric load forecasting. *Journal of Intelligent and Robotic Systems. 31*, 7-68. https://doi.org/10.1023/A:1012402930055

[24] Groß Arne, et al. (2021) Comparison of short-term electrical load forecasting methods for different building types. *Energy Informatics. 4(Suppl 3)*, 13. https://doi.org/10.1186/s42162-021-00172-6

[25] Son Namrye (2021) Comparison of the deep learning performance for short-term power load forecasting. *Sustainability. 13(22)*, 12493 https://doi.org/10.3390/su132212493

[26] Ribeiro Andrea Maria NC, et al. (2022) Short-and very short-term firm-level load forecasting for warehouses: A comparison of machine learning and deep learning models. *Energies. 15(3)*, 750 https://doi.org/10.3390/en15030750

[27] Guo Weilin, Liang Che, Mohammad Shahidehpour, & Xin Wan (2021) Machine-Learning based methods in short-term load forecasting. *The Electricity Journal. 34(1)*, 106884 https://doi.org/10.1016/j.tej.2020.106884

[28] Mohammad Faisal, Ki Boem Lee, & Young-Chon Kim (2018) Short term load forecasting using deep neural networks. *arXiv preprint arXiv:1811.03242*,

[29] Din Ghulam Mohi Ud & Angelos K Marnerides (2017) Short term power load forecasting using deep neural networks. *2017 International conference on computing, networking and communications (ICNC)*, 594-598 https://doi.org/10.1109/ICCNC.2017.7876196

[30] Kraskov Alexander, Harald Stögbauer, & Peter Grassberger (2004) Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 69(6)*, 066138. https://doi.org/10.1103/PhysRevE.69.066138

[31] Zhang Rui, et al. (2013) Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine. *IET Generation, Transmission & Distribution. 7(4)*, 391-397 https://doi.org/10.1049/iet-gtd.2012.0541

[32] Hossen Tareq, Arun Sukumaran Nair, Radhakrishnan Angamuthu Chinnathambi, & Prakash Ranganathan (2018) Residential load forecasting using deep neural networks (DNN). *2018 North American power symposium (NAPS)*, 1-5. https://doi.org/10.1109/NAPS.2018.8600549

[33] Cai Qiuna, et al. (2020) Short-term load forecasting method based on deep neural network with sample weights. *International Transactions on Electrical Energy Systems. 30(5)*, e12340. https://doi.org/10.1002/2050-7038.12340

[34] Goodfellow Ian, Yoshua Bengio, Aaron Courville, & Yoshua Bengio.(2016). *Deep learning*. MIT press Cambridge.

[35] Simonyan Karen & Andrew Zisserman (2014) Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*,

[36] He Kaiming, Xiangyu Zhang, Shaoqing Ren, & Jian Sun (2016) Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778 https://doi.org/10.1109/CVPR.2016.90

[37] Ronneberger Olaf, Philipp Fischer, & Thomas Brox (2015) U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[38] Jadon Aryan, Avinash Patil, & Shruti Jadon (2024) A comprehensive survey of regression-based loss functions for time series forecasting. *International Conference on Data Management, Analytics & Innovation*, 117-147. https://doi.org/10.1007/978-981-97-3245-6_9

[39] Masood Zaki, Rahma Gantassi, & Yonghoon Choi (2024) Enhancing short-term electric load forecasting for households using quantile LSTM and clustering-based probabilistic approach. *IEEE Access. 12*, 77257-77268. https://doi.org/10.1109/ACCESS.2024.3406439

[40] Islam Md Tariqul, MJ Hossain, & Md Ahasan Habib (2025) Data-Driven Dynamic Optimization for Hosting Capacity Forecasting in Low-Voltage Grids. *Energies. 18(15)*, 3955. https://doi.org/10.3390/en18153955

[41] Aburaed Nour, et al. (2022) A comparative study of loss functions for hyperspectral sisr. *2022 30th European Signal Processing Conference (EUSIPCO)*, 484-487. https://doi.org/10.23919/EUSIPCO55093.2022.9909827

[42] Cao Zhaojing, et al. (2019) Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting. *IEEE Transactions on Power Systems. 35(3)*, 1881-1897. https://doi.org/10.1109/TPWRS.2019.2946701

[43] Pinheiro Marco G, Sara C Madeira, & Alexandre P Francisco (2023) Short-term electricity load forecasting—A systematic approach from system level to secondary substations. *Applied Energy. 332*, 120493. https://doi.org/10.1016/j.apenergy.2022.120493

[44] Cui Ziti, et al. (2024) High-precision identification and prediction of low-voltage load characteristics in smart grids based on hybrid deep learning framework. *International Journal of Low-Carbon Technologies. 19*, 2656-2666. https://doi.org/10.1093/ijlct/ctae221

[45] Ali El-Sayed Ali Hussien, MH Alham, & Doaa Khalil Ibrahim (2024) Big data resolving using Apache Spark for load forecasting and demand response in smart grid: a case study of Low Carbon London Project. *Journal of Big Data. 11(1)*, 59. https://doi.org/10.1186/s40537-024-00909-6