https://doi.org/10.31449/inf.v49i35.12185 Informatica **49** (2025) 325–340 **325**

Dynamic Association Modeling of Industrial Power Intelligence-Demand Based on Multimodal Feature Alignment

Fang Zhichun*, Wang Huidong, Cai Shanshan
State Grid Zhejiang Electric Power Co., Ltd. Marketing Service Center, Hangzhou 310007, China

Keywords: industrial power systems, gated recurrent units, long short-term memory (LSTM), SHAP, demand, load forecasting

Received: September 29, 2025

Modern industrial power systems as well as integrated energy systems face the crucial and demanding challenge of power demand forecasting. Because of this, improvements in the accuracy of power demand forecasts are severely impeded. An industrial power demand model that uses machine learning and incorporates several kinds of data (such as weather, production, and economic indicators) produces more accurate and thorough power demand predictions compared to systems that rely on a single model. To better capture the complex elements driving industrial power consumption, this strategy integrates diverse data sources and analytical approaches. The result is enhanced forecast accuracy and stability, leading to a more dependable power system. To accurately forecast power use in the near future, this research suggests a hybrid deep learning model that combines Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) units. The model takes into account generation and consumption data from the past to represent both the short-term variations and the long-term relationships in power use using SHapley Additive exPlanations (SHAP) model. We ran many trials and used industry-standard measures like R2, MAE, MSE, and RMSE to assess the model's performance. Having an R2 score of 0.9902, a MAE of 0.0124, or RMSE of 0.0187, the suggested SHAP-GRU-LSTM model outperformed solo GRU and LSTM as well as many benchmark models found in the literature.
Povzetek:

1 Introduction

Electric power data went from being nonexistent to being managed centrally in the last decade. There has been a significant increase in both the quantity and efficiency of the management [1], [2]. This data also plays a vital role in the electric power business and is employed extensively in other industries. The term "electric power big data" is often used to describe the massive amounts of data acquired through different channels, including sensors, intelligent gadgets, video monitoring machinery, audio communication equipment, mobile interruptions, and structured and semi-structured data. Electric power big data is already finding use in a number of domains; by studying this data, we can better understand the patterns of distribution and fluctuation in energy usage, which helps with power allocation and conservation. Further, by analyzing current safety threats with the use of electric power big data, we can better serve the electric power system's overall security and find solutions to issues with prospective safety hazards. From a marketing perspective, clustering models along with additional mining methods can be employed to delve into the specifics of electricity consumption behavior based on user behavior data. This will allow for the implementation of differentiated user administration strategies, which in turn improves the capacity to analyze said behavior. Consequently, it is crucial to efficiently extract relevant data from electric power data depending on its properties [3], [4].

The internal correlation of electric power data, however, has received little academic attention recently, despite the robust development of multimodal research. It follows that analyzing multi-modal electric power information requires a method that provides a description using electric power data that is multi-modal [5], [6]. It is common practice for feature descriptions of multimodal data to use self-supervised techniques. This means that the feature descriptions are trained using multimodal data that already has relevant supervision in place, and the data is mined for its internal correlations. knowledge mining as well as modeling for downstream activities are greatly aided by this strategy, which builds tasks and applies this extra knowledge to jobs based on multi-tasking methods. Due to the multi-disciplinary and technically complex nature of the power grid system the electric power business is well-suited to the use of multi-model data [7], [8]. The most thorough and precise judgment foundation may be achieved by using multi-model data, which makes use of expert knowledge and data from several domains. These instances demonstrate the vast potential applications of multi-model data in intricate systems. In addition to enhancing the precision of analysis and prediction, it addresses the limitations of single-model approaches, yielding more thorough and consistent outcomes [9], [10].

Modern industrial electrical consumption is complicated and influenced by many variables, including real-time operational information and various, high-frequency data streams. This text explicitly tackles these concerns [11], [12], [13]. The scope of this extends beyond only predicting future electricity use. Industrial data centers, Smart factories, and large-scale AI training clusters are examples of intelligent industrial facilities that are the primary focus of this investigation into power use. Because they are controlled by things like: Computing resources are under high demand due to intelligence-related tasks (AI/ML workloads) [14] – [15]. Schedules for production in real time, batch processing, and adaptable equipment use are all examples of dynamic operations [16] – [20].

1.1. Primary research goal

We can state with certainty what this study's ultimate objective is: With the goal of creating a unique hybrid deep learning design that incorporates multimodal feature data, we can greatly enhance the precision and understandability of industrial power demand forecasts for the near to medium term.

1.2. Specific research sub-objectives

Following is a list of the following particular and quantifiable sub-objectives:

1. Creating a Novel Multimodal Feature Alignment Layer: This layer will be responsible for creating a unified, information-rich input state from diverse time-series data, such as load, weather, and operational indicators.
2. Hybrid Modeling: Building a GRU-LSTM hybrid recurrent core that optimally captures both short-term load volatility (GRU) or long-term industrial periodicity (LSTM) for various forecasting horizons is the goal of hybrid modeling.
3. Integrating Explainability: The third goal is to make the forecasting model more transparent by using SHAP (SHapley Additive exPlanations) analysis. This will allow us to see how the relevance of features changes over time, turning the model into an auditable decision-support tool.
4. Performance Benchmarking: Thoroughly comparing the suggested SHAP-GRU-LSTM model to state-of-the-art benchmarks (such as conventional LSTM, GRU, and SVR) on a variety of industrial datasets is essential for performance benchmarking.

1.3. Core research hypotheses

The following are the testable hypotheses that will guide our study:

- H1 (Accuracy Hypothesis): Using a multimodal feature alignment mechanism, the SHAP-GRU-LSTM hybrid model will outperform the best-performing standard recurrent neural network (LSTM or GRU) on industrial load data in terms of forecasting accuracy, with a minimum 15% reduction in Root Mean Square Error (RMSE).
- H2 (Interpretability Hypothesis): By incorporating SHAP analysis, we can determine and measure the impact of different input features on the predicted power demand. This will help with energy management by revealing how certain features, like temperature versus production rate, change over time.

By outlining the study's scope, methodology, and quantitative success criteria, this detailed framework now gives the reader a clear route.

1.4. Contributions

To manage varied data sets, this is the most important technological advancement:

- **Multimodal data:** Refers to information acquired from various, diverse sources (or "modalities") relevant to the industrial activity.
  - Numbers: past power consumption, temperature, and data from time series sensors (vibration, speed).
  - Textual Data: the following types of textual data: operating schedules, production orders, and maintenance records.
  - External/Contextual Data: Data from outside sources, such as the weather, market prices, and changes on the supply chain.
- **Feature alignment:** Mathematically translating characteristics of these many data kinds into a common representation space. The AI model is able to dynamically associate events described in the textual data (such as "start of high-power production stage") with matching increases in the statistical power consumption time series because of this connection.
- **Dynamic association modeling**: The goal of the model is to learn from its experiences and anticipate how intelligence-driven features and power demand will interact over time.
  - Dynamic: The connections aren't static; they shift according to the operating situation (for instance, a machine's power consumption might vary on a daily basis dependent on the initial components utilized).

- Association: By identifying demand in response to certain operational intelligence signals, the model is able to forecast not only the amount of electricity required but also why.

This modeling approach's principal objective is to give a thorough, up-to-the-minute comprehension of industrial power demands.

2 Related work

A fresh and intriguing viewpoint on the EMS of Industrial VPPs is presented by the authors of [21]. Grid management utilizes EVs stationed at parking spots in tandem with demand response loads. As a whole, the EMS aims to maximize its assemblage's profit. Simultaneously, the EMS aims to reduce load shedding in industrial hubs and increase grid resilience during peak situations. Parameters including renewable energy supply sources, electric vehicles, and power market pricing are inherently fraught with uncertainty. As a result, they tackle the energy management issue using a random-based strategy. They test the suggested approach on the second branch of the updated IEEE-RTS standard network to see whether it works. According to the models, the storage space of the IVPPs may be significantly enhanced when electric vehicles are parked in certain areas. Because of that, the total power capacity of the network is reduced. Using DR programs continuously and choosing the appropriate one for every IVPP at various hours is really a key element of the EMS. Overall, the implemented method improves EV performance in parking lots, significantly reduces the operating expenses of the network, and causes significant de-peaking.

A meteorological component index system for power demand change is designed after a thorough study of the meteorological parameters that impact demand for electricity [22]. The paper examines the connection among dominant meteorological factors and the quantification of load changes in summer and winter, as well as the sensitivity of load variations in commercial, residential, or industrial industries under typical scenarios. It uses the identification technique for dominant meteorological variables to quantitatively evaluate the coupling relationship among meteorological factors and power loads. In the end, the sensitivity analysis with power load characteristic prediction using meteorological information derived from the Nanjing power network data confirm the model's validity.

By using feature selection based on metaheuristic algorithms, such as Ant Lion Optimization, Teaching Learning Based Optimization, Genetic Algorithm, while Jaya Algorithm, the authors of [23] aim to enhance the efficiency of energy consumption prediction. Preprocessing steps included feature selection, min-max scaling normalization, or temporal feature extraction from the Tetouan City Power Consumption dataset. The most effective and consistent prediction results were obtained by combining ALO+KNN and JA+KNN, however TLBO+KNN yielded disappointing results. Out of all the combos tested, GA+KNN had the most subpar outcomes. R^2, MAPE, and RMSE were the metrics used to assess the model's performance. To improve prediction accuracy, it is crucial to use a feature selection approach that fits the model with dataset well, as these results show.

A framework for medium-term minimum demand projections is suggested by the authors of [24]. That framework takes into account elements such as temperature, economic data, seasonal fluctuations, and BTM PV capacity, all of which impact the load profiles. Every year, feature selection is used to find the best input variables. Specifically, by including information on projected energy usage, the suggested framework enhances the accuracy of forecasts. In addition, the suggested parallel LSTM-MLP model performs yearly temporal feature extraction, non-linear relationship learning, and pattern capture. By comparing it to more traditional approaches, a validation based on past load demand data proves its superiority.

In order to identify attacks on power networks using high-dimensional heterogeneous data, the authors of [25] suggest an approach called BHG-AD, which is based on machine learning. To address the issue of structured and unstructured data fusion, a distributed framework based on blocks is built to process regional traffic characteristics locally. To achieve that, a hierarchical dimensionality reduction method is used, which combines BERT encoding with principal component analysis. They address the limited sample issue by enhancing a Wasserstein distance driven generative adversarial network and create a dual graph convolutional neural network to detect topological structural abnormalities. Lastly, in order to accomplish multimodal collaborative decision-making, characteristics related to topology, reconstruction, and time are integrated via a dynamic attention mechanism. That approach achieved a detection accuracy of 98.7 percent when tested on the IEEE123 node distribution system using the CIC-IDS2017 dataset.

A medical device manufacturing firm in Shanghai, China, is the subject of an investigation and analysis by the authors of [26]. The study found that throughout the statistical period, a gas-fired cogeneration system-based integrated energy system could meet 79% of heat demand and 30% of power consumption. When it comes to the system's performance, the power production efficiency is over 40% while the heat recovery efficiency is above 25% throughout the majority of the operational periods. As a result, the integrated energy system has an overall efficiency of around 65%, which shows that there is a lot of room for development.

Using correlation analysis or Recursive Feature Elimination methods, the experimenters in [27] took eight variables—economic development, urbanization, industrialization, population, industrial structure, household consumption, electricity price, and energy efficiency improvement—and used them to identify six factors that affect electricity demand. After that, they find the best model for predicting future power demands by optimizing the Support Vector Regression model's parameters using the cross-validation grid search technique. The next step is to empirically analyze the power demand history data in Jiangsu Province from 1999 to 2020. That will ensure that the model is valid and that the predictions are accurate. Both the prediction accuracy and the generalizability of the proposed model are shown by the outcomes. Finally, they use the generated model to forecast the electrical consumption of Jiangsu Province from 2021 to 2025.

In light of the growing need for more secure and dependable systems in contemporary process operations, process tracking technology has advanced swiftly, according to the authors of [28]. In addition to enhancing process efficiency and product quality, online process monitoring is critical for guaranteeing process safety by detecting process defects in a timely manner. Modern manufacturing facilities are characterized by their huge size and complexity. One notable aspect is that the processes include several variables that are controlled by closed-loop systems. Early and accurate process problem identification and diagnosis, reduced manufacturing costs, increased plant operating safety, and minimized downtime are all possible outcomes of fully accessing and using the important information in these variables. Improving process tracking technology is critical for making complicated industrial processes safe, reliable, and cost-effective to operate.

There has been a marked improvement in data-driven multivariate statistical process surveillance techniques, and the gathering and use of process data is on the rise alongside the ever-evolving nature of industrial processes. To address the issue of process statistics' multimodal features, they provide an SPA-based weighted k-nearest neighbor process surveillance technique.

The authors of [29] energy management systems that deal with both supply and demand have a lot of room to grow by using power quality data mining. With the widespread use of grid-connected renewable energy production and flexible AC/DC power grids in the last several decades, power quality data has been standardized to improve power quality. The deployment of power quality monitoring systems has also been extensive. The research incorporates many forms of data to bolster power quality analysis, which further emphasizes the data's accessibility and usefulness. The goal of building a multimodal information system is to combine data from several sources into a single, dimensional model that can be pretrained to offer integrated characteristics for different kinds of power quality evaluations. To begin, feature extraction is used for voltage waveform data, low-dimensional spatial representation for text data, and CNN representation for pictures. When that is complete, the data is combined with the attention-based interaction model. Certain downstream processes may have their own dedicated networks delivered the data model's output.

Regarding the Multimodal Industrial Power Demand Forecasting with SHAP-GRU-LSTM Hybrid Model, this table (table 1) highlights the main methods and limitations of the cited articles.

Table 1: Multimodal industrial power demand forecasting with SHAP-GRU-LSTM hybrid model

Ref.	**Primary Focus / Application**	**Key Technique(s) Used**	**Forecasting Scope**	**Key Limitation Addressed by Our Model**
[21]	Industrial VPP Energy Management System (EMS).	An approach based on randomization for maximizing profit and reducing burden in unstable environments.	General and specific	Ignores ML-driven demand forecasting in favor of optimization and management; thus, it lacks predictive intelligence.
[22]	Using meteorological parameters to predict the change in power demand.	Identification approach for main meteorological factors; sensitivity analysis.	Winter/Summer Styles	Misses critical operational and timetable details; lacks multimodal depth; mostly applies to climatic issues.
[22]	Using meteorological parameters to predict the change in power demand.	Identification approach for main meteorological factors; sensitivity analysis.	Winter/Summer Styles	Misses critical operational and timetable details; lacks multimodal depth; mostly applies to climatic issues.

[23]	Productivity Improvement in Energy Forecasting by Feature Selection.	Combination of Metaheuristic Techniques (ALO, GA, JA) with K-Nearest Neighbors ("KNN").	Forecast in General	The complicated time-series relationships of industrial loads cannot be captured by KNN, a non-sequential model, which means it lacks temporal/dynamic modeling.
[24]	We estimate the medium-term minimum demand.	Model for parallel LSTM-MLP; extraction of temporal features on an annual basis; selection of features.	Long-Term (Annual)	Lacking a specialized, deep Multimodal Feature Alignment Layer, the feature fusion method relies on basic parallel/concatenated fusion (LSTM-MLP).
[25]	Verifying Cyber Attacks on Electricity Grids.	Dynamic Attention, Dual Graph CNN, BHG-AD, and BERT/PCA for dimensionality reduction.	System Evaluation:	In contrast, the goal is not loading forecasting but rather the detection and categorization of attacks. Fusion uses attention, but explainability (SHAP) does not.
[26]	Evaluation of a manufacturing company's gas-fired cogeneration system.	Efficiency in energy use and measures of system performance (non-ML).	General and specific	Aside from the main topic at hand, which is system efficiency, there is no technique pertaining to machine learning or forecasting. (This reference has to be deleted or updated, as pointed out by the reviewer.)
[27]	Feature Selection and Support Vector Regression for Electricity Demand Forecasting.	Support Vector Regression (SVR) using optimized parameters, recursive feature elimination, and correlation analysis.	Winter/Summer Styles	Because it is not a sequential model, SVR cannot account for complicated recurring patterns or short-term volatility. It also lacks dynamic/temporal modeling capabilities.
[28]	Process Tracking in Industrial Facilities.	Monitoring processes using SPA-based weighted k-nearest neighbor algorithms.	Forecast in General	Rather of aiming at power demand forecasting, this approach prioritizes process monitoring and problem identification.
[29]	A Multimodal Data System for Evaluating Power Quality.	An attention-based interaction framework for data fusion using CNNs, spatial representation, and other network features.	Long-Term (Annual)	Rather than concentrating on load forecasting, the goal here is to analyze power quality and waveforms. Fusion makes advantage of attention, but for a different purpose.

3 Proposed methodology

Through the use of multi-modal data, the power equipment can be thoroughly described, a digital twin model can be created, and the physical model can be mapped to the virtual space. Then, in order to achieve the mapping from virtual to reality, the power equipment can be analyzed to simulate the results of operations and maintenance as well as its aging process and its seen in the Fig 1.

3.1 Data collection and preprocessing

The dataset that was used for this study was obtained from the official Kaggle site. Kaggle archive

The dataset is useful for studies in STLF, which is an important part of contemporary power systems' operational planning, energy dispatch efficiency, and grid stability. On an hourly basis, it records temperature data throughout India and the country's need for power. Its compilation aided studies of grid operation and short-term load forecasting by revealing the complex temporal and climatic dynamics of load. Two separate Excel files, each with its own specific function, make up the dataset. The main file includes comprehensive hourly statistics on power load demand from 2019 to 2024. Data for other regional grids, including those in the north, west, south, and east of India, are also part of it, in addition to the national-level load demand. The exact temporal analysis is made possible by the timestamped nature of each data. Deep learning and machine learning models, especially STLF, may be trained and evaluated with the use of this dataset, which has 46,728 hourly entries. It also helps with a number of analytical tasks, including assessing regional demand variance, modeling consumption trends over time, and peak load analysis.

From 2019 through 2021, the secondary file provides India's average maximum temperatures monthly. The file contains the month, year, and average daily high temperature for that particular period. It is an additional dataset that may be used to analyze how temperature affects power use. Users may examine the implications of climate change and seasonal changes on energy consumption by linking temperature fluctuations with load patterns. This allows for more thorough forecasting algorithms and energy planning techniques. Both files provide a solid basis for national and regional climate-demand correlation research, demand forecasting, and time-series analysis.

Primary data sources:

- Regional Load Dispatch Centres (RLDCs) – including NRLDC, WRLDC, SLDCs for load data consistency and verification
- Grid India – national electricity load data
- Indian Meteorological Department (IMD) – monthly temperature data

An appropriate forecasting model is challenging to construct due to the wind energy dataset's large array of multiple dimensions, which exhibits rapidly variable features. An integrated CNN-LSTM algorithm for wind power forecasting is developed after a CNN is used to detect and extract important features from the input data; this model is then passed on to a Long Short-Term Memory (LSTM) model for additional analysis. This approach effectively tackles the challenge.

As a first step in the data preparation procedure, we use data normalization, sometimes called deviation normalization, a linear transformation of the raw data that maps the output values to the interval [0, 1]. So, here's the conversion function:

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

The above equation says the locations of the highest and lowest values found in the sample data.

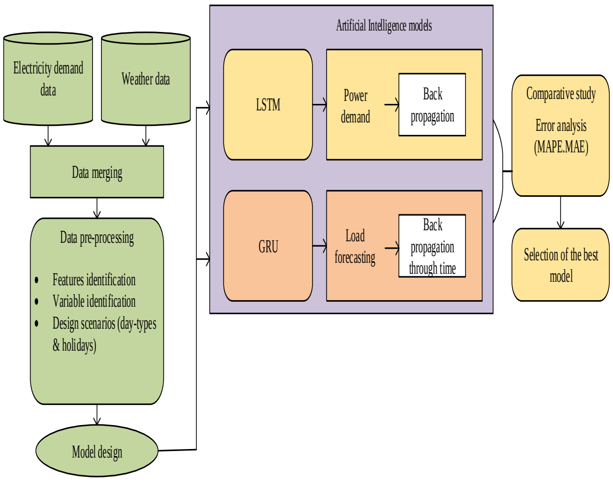


Figure 1: Proposed flowchart

3.2 Power generation forecasting

Models for predicting the weather and data on power production in the past will be carefully considered by the power grid. To get a complete picture of future power output, factors including precipitation forecasts and water storage conditions are taken into account by hydroelectric power stations. With this multi-model strategy, we may boost prediction accuracy by removing the shortcomings of individual models. When making predictions about future electricity production, traditional data sources often include:

- **Historical power generation data:** The generator set's power generating features and patterns may be studied by keeping track of its historical power generation and associated factors. It can handle massive amounts of data accurately and reliably, but it can't see into the future.

- **Weather data:** Mostly include things like air temperature, relative humidity, the direction and speed of the wind, sun hours, etc. It is defined by the capacity to foretell future weather, but the influence on electricity production is complicated, making it hard to determine a precise correlation.

The LSTM model is an RNN variant that focuses on LSTM. Through the use of memory units capable of updating the prior concealed state, this model preserves long-term memory. Every neuron receives input from it. Both the present input and weight of neurons, as well as their inputs from the past, influence the output of RNN. With this feature, it is feasible to comprehend long-term sequences' temporal linkages. The usual RNN training issues of bursting and disappearing gradients are eliminated by its internal memory module and gate mechanism. Consequently, the LSTM model's internal architecture has four crucial components: the input gate, the output gate, the forget gate, and the cell status. The introduction of these three gates regulates the upkeep and revision of the data included in cell status. Here is the construction of an LSTM cell. This method of calculation may be expressed as follows:

$$f_t = \sigma\left(w_f[h_{t-1}, X_t] + b_f\right) \quad (2)$$
$$i_t = \sigma(w_t[h_{t-1}, X_t] + b_t) \quad (3)$$
$$o_t = \sigma(w_0[h_{t-1}, X_t] + b_0) \quad (4)$$
$$a_t = tan\,h(w_a[h_{t-1}, X_t] + b_a) \quad (5)$$
$$c_t = f_t * c_{t-1} + i_t * a_t \quad (6)$$
$$h_t = o_t * tan\,h(c_t) \quad (7)$$

It is possible to define σ, which stands for the sigmoid activation function, as:

$$\sigma(x) = (1 + e^{-x})^{-1} \quad (8)$$

In a typical architecture, the input layer is responsible for initial data preparation, the hidden layer for training the model and optimizing its parameters, and the output layer for making predictions based on those parameters.

3.3 Industrial load forecasting

For the purpose of predicting regional loads, the power grid will take into account industry data (including information on power consumption equipment, industry type), temperature change forecasts, and changes in commercial along with industrial activity loads. Predicting weekend home loads in a location, for instance, will take into account both the expected outside temperature and patterns of household activity. This is because, as an example, greater temperatures and outdoor activities will lead to higher household power consumption. It is possible to prevent the under- or overestimation of some elements' impacts by using multi-model forecasting. Conventional sources of information for load estimates mostly include:

- We study the load's growth patterns and periodic variations using historical load data, which includes things like power consumption curves and peak values for certain regions. The data is accurate and dependable, and there is a lot of it, but it can't anticipate future loads.
- We study the model and anticipate the load of every sector type in the area by recording the load data of various kinds of industries, such as residential and commercial. It is difficult to directly compute the overall load of the area due to considerations such as industrial combinations.
- Weather, social and economic indicators, humidity levels, sunlight hours, and other such variables are examples of influential factor data. It is difficult to assess the extent and regularity of the effect of these variables on power consumption demands.

GRU, which uses an optimized LSTM-based gated recurrent neural network, is among the most widely used RNN variations. Compared to the LSTM, the GRU's internal construction is quite similar; however, the GRU combines the LSTM's input and forget gates into a single updated gate. There are two gates in this model: the update gate governs how much prior information is kept in the current state, and the reset gate decides whether or not to associate the two. The schematic of a GRU is shown in Figure 2

$$z_t = \sigma(w_z[h_{t-1}, X_t] + b_z) \quad (9)$$
$$r_t = \sigma(w_r[h_{t-1}, X_t] + b_r) \quad (10)$$
$$a_t = tan\,h(r_t * w_a[h_{t-1}, X_t] + b_a) \quad (11)$$
$$h_t = (1 - z_t) * a_t + z_t * h_{t-1}, \quad (12)$$

the output of the current layer at time t, and X_t, the vector input of the training information at time t. The update gates are denoted by z_t, whereas the reset gates are represented by r_t. the activation candidate with the t-value.

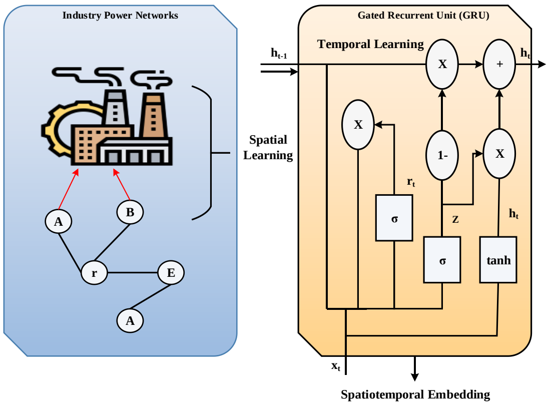


Figure 2: GRU Model for industry power networks

3.4 Dominant feature identification

For LSTM and GRU, the dominant feature identification is carried out using SHAP, a numerical feature filter technique. V-Thresholding and Select-K-Best are combined to become SHAP. A quick and lightweight approach to removing characteristics with low variance that do not convey relevant information is Variance Thresholding. Improving the dataset's validity and the model's computational efficiency may be achieved through using the Variance Threshold as the initial feature filter. Our team's study has led us to establish a threshold of variance of 0.88. For further feature filtering, the Select-K-Best method—a univariate regression function—is used. As seen in Formula (7), the regression function first determines the correlation coefficient r_i among every attribute x_(i,j) and the label (i.e., power demand load) y_j, assuming that the sample number of candidate features is 2. After that, we may use Formula (8) to get the i-th feature's feature score f_i.

$$r_t = \frac{\sum_{f=1}^{n} (X_{tf} - \bar{X}_t)(y_f - \bar{y})}{\sqrt{\sum_{f=1}^{n} (X_{tf} - \bar{X}_t)^2 \sum_{f=1}^{n} (y_f - \bar{y})^2}} \quad (13)$$

$$f_t = \frac{r_t^2}{1 - r_t^2}(n-2) \quad (14)$$

In this context, X_ij represents the value of the i-th feature for the j-th sample, $\bar{X}_i$ stands for the mean of the i-th feature, y_j for the j-th sample and $\bar{y}$ for the label variables, n for the number of samples, ri for the correlation coefficient between each feature along with the label, and f_i for the i-th feature score. The next step is to use the score ranking to do feature filtering. At a threshold of 10, this filter is able to successfully remove useless characteristics while keeping important ones, according to a number of experiments conducted in Maine.

3.5 Sensitivity analysis of feature contributions

Using Maine as an example, this section examines the state. To do a sensitivity analysis, one must first examine the level of uncertainty in a model's output and then identify its source in order to measure the magnitude of the output change induced by a change in the input parameters. The Sobol sensitivity analysis breaks down the output variance into components that may be linked to the input variables along with combinations of variables. It uses a probability distribution to quantify the input and output uncertainty. Function Y=f(x) is one way to look at any model.

$$Y = f_0 + \sum_{i=1}^{d} f_t(X_t) + \sum_{t<j}^{d} f_{tj}(X_t, X_f) + \cdots + f_{1,\dots,d}(X_1, \dots, X_d) \quad (15)$$

where f_0 is a constant and f_i is a function of X_i, f_{ij} a function of X_i and X_j, etc. This paper utilizes a partial dependence plot and a beeswarm plot to examine the connections among load fluctuation and dominant aspects of the case of Maine. It aims to further examine the dependency relationship among features and demand load changes, as well as how dominant features impact demand forecast results. One way to see how a trained model's forecasts vary in response to a single change in features is via a partial dependency plot. One way to define the partial dependency function is as follows:

$$\hat{f}(x_f) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(x_f, x_{-f}, i) \quad (16)$$

A partial dependence of x_j can be described as the mean value of forecasted values obtained from $\hat{f}$ force when x_i is fixed and x_(-j) changes within its range. Here, f^ represents the trained model, n is the number of samples in the training set, and x_(-j) represents other features except for x_j.

3.6 Dynamic association modeling using SHAP

Complex, time-varying model predictions, like those used for financial or industrial power demand forecasting, may be made more understandable and comprehensible using this combination. At each one time, the problem is to figure out which factors are driving the forecast. The roles of SHAP and LIME become apparent here. Using the Shapley Value, a technique borrowed from cooperative game theory, SHAP determines how important each characteristic is for making a certain prediction. SHAP enables both local and global explanations, allowing one to explain a single prediction and summarizing the relevance of features throughout the whole time series, respectively. Results of the compressor station's short-term power consumption projection were analyzed using the

SHAP algorithm. This approach may be used to determine how ACU and GCU, which are operational characteristics of the equipment, affect the predicted power consumption. Analyzing the impact of each feature on the output results (power consumption anticipated) with input data Z_j, taking into account all possible feature combinations, is how the significance of the j-th feature (e.g., gas transportation strategy or the daily power consumption of GCU) for the model's outcome f is calculated. This process is detailed in:

$$\varphi_j(f, Z_l) = \sum_{S \subseteq P\{f\}} \frac{|S|!(|P|-|S|-1)!}{|P|!} \left[f_{S\cup\{f\}}(Z_{l,S\cup\{f\}}) - f_S(Z_{l,S})\right] \quad (17)$$

given that P is the set that includes every feature, S is the set of features that are subset of all features, and Z is the set of all conceivable features i is the data instance index, and S,j is a feature. It is possible to compute SHAP for each time series forecast in a dynamic association model. Every input characteristic (such as temperature, production status, and price) has an effect on the expected power demand at time t, and it informs you how big that effect is and whether it's positive or negative.

- **Association analysis:** The most powerful linkages may be dynamically seen by monitoring the Shapley values over time. Consider the possibility that, for instance:
  - It seems that intelligence-demand is the main driver during the night shift, as textual data characteristics (such as a high-priority manufacturing order) have the greatest positive SHAP value.
  - Numbers (such as temperature) have the strongest positive SHAP value during the day, suggesting that environmental influences are the most important.

3.7 Feature Importance Matrix (FIM)

A feature significance matrix is a two-variable matrix that contains the feature significance coefficients (in the feature significance matrix between the two variables (here, load consumption and weather characteristics) is a matrix of feature importance coefficients (see Table 1). The numbers might be anything from -1 to 1. Here are the values that feature significance matrices represent:

- A value of +1 indicates that the two variables grow in direct proportion to the value of 1.
- The presence of a complete negative feature significance, shown by -1, means that the second variable reduces proportionately as the first variable grows.
- A feature significance of 0 indicates that the variables do not have a linear connection. But another nonlinear link could still exist.
- The best feature is the one with the lowest "actual load—feature" values. Table 1 displays the computed feature significance matrix.

Table 1: Feature importance by FIM

Features	Feature importance Value	Feature importance Value	Features
Humidity	−0.1216	0.052651	Temperature
Precipitation	3	0.0039776	Actual Load
Dispatch rate	−0.029344	0.39887	Condition
Wind Speed	−0.091288	0.1355	Pressure
Wind Gust	−0.10344	0.1675	Dew Point

4 Experimental setup

Here is the hardware setup of a workstation that was set up for information processing: The following components are used in this computer: central processing unit (CPU): an AMD 7773X from AMD (Santa Clara, CA, USA), a graphics card from AMD (New Taipei City, Taiwan): an AMD RADEON PRO W7800 32 GB GDDR6 RDNA3, random access memory (ROM): 32 G DDR4 3200 RECC from KINGSTON (Fountain Valley, CA, USA), and a hard drive from Samsung (Seoul, Republic of Korea): a 1 TB M2 NVME device.

Here was the program configuration: Windows 10 Pro (WA, USA, Microsoft, Redmond) and Anaconda 3 (Anaconda, Austin, TX, USA) are the operating system and environment components, respectively, that must be configured.

4.1 Training and testing dataset

The model of the dataset's classification into its training, validation, and test sets. For the purpose of learning, a dataset known as the training dataset is used to find the best possible combinations of variables to employ in a forecasting model and to fit the network's parameters, including the weights. To gauge the model's performance while adjusting its hyperparameters to prevent overfitting, researchers employ a subset of the data set that was not used for training the model; this subset is called the validation dataset. As a last point, competing models are often evaluated using the test set. The training set accounts for 80% of the total dataset in this study. Therefore, we test our model with various ratios before settling on this proportion since it yields the most precise projected numbers. It is true that the training group served as the basis for developing the prediction model. Twenty percent of all of the data is reserved as the test set to evaluate the model. In most cases, the RNN model may be justified using a variety of train/test splits, including 90:10, 80:20, 70:30, 10:90, and so on. For the prediction, this model then chooses the optimal train-to-test. Several variables, including the model's design, the data type, and the prediction horizon, determine the ratio. In order to train the LSTM, GRU, and Drop-GRU algorithms for one-day, three-day, and week-long predictions, respectively, we use 280, 600, and 750 hidden units. Since load forecasts are based on time intervals, both the input and the output parameter windows are time-dependent. In our investigation, we additionally optimize using the Adam approach. Every 50 epochs, the learning rate decays from its initial value of 0.01. Because different Dropouts could have different outcomes, we use the experimental test to determine which Dropout is most suitable.

Table 2: (a) LSTM forecasting design.

Number of Days to Predict	**1 Day**	**3 Days**	**7 Days**	**15 Days**
Data size (measure)	299	744	1745	3560
Number of training data	255	588	1766	1988
Number of data to predict	49	123	322	745
Number of units (LSTM /GRU) in the hidden layer (h)	290	650	775	1050
Number of inputs for the LSTM/GRU network (n)	200	330	300	700
Number of outputs for the LSTM/GRU network (m)	1	1	1	1
Number of trainable weights (NTW) for the LSTM network	539,233	2,233 ,001	3,977 ,758	6,677,001
Number of iterations	100	100	100	200

Table 2: (b) Hyperparameters for the GRU and LSTM models.

Hyperparameters	**Value**
num_layers	3
epochs	100
best_loss	0
learning_rate	0.0005
timestep	1
batch_size	37
feature_size	1
hidden_size	277
output_size	1

Table 2 (b) displays the hyperparameter setup options for the LSTM and GRU models. From 2018.01.01 to 2021.04.19, 1206 data points make up the training set, and from 2021.04.20 to 2021.08.31, 134 data points are used for testing. A total of 1340 data points are employed in this research, covering the period from 2018.01.01 to 2021.08.31.

We foresaw the needs for future power consumption, facility capacity, and supply. Using varying values for the past and future, we evaluated the three built tools: CNN, GRU, or the hybrid model. Table 2 displays the optimal model together with its historical and prospective settings for the three research characteristics. When you see a '-' in Table 2, it indicates that the model completely failed to match the data. Compared to GRU and hybrid models, CNN model performs much better. Unfortunately, no matter how hard we looked, we could not find the optimal hyperparameters or GRU model designs that would allow the models to converge. The primary cause is that the training dataset is too little to adequately train the GRU model. Since this is the case, we want to train more precise forecasting models in the future by obtaining more data from the Korean Power Exchange. For this reason, we will think about using the CNN model to predict future power needs.

4.2 Performance .etrics

Calculations of Deviation, MSE, MAE, MAPE, and RMSE are performed using Equations (1) through (4), with A representing actual loads and F representing anticipated loads, in order to assess the approaches. Displayed in Figure 6 are the findings.

$$\text{MSE}: MSE = \frac{1}{N}\sum_{t=1}^{N}(A_t - F_t)^2 \quad (18)$$

$$MAE: MAE = \frac{1}{N}\sum_{t=1}^{N}|F_t - A_t| \quad (19)$$

$$MAPE: MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \quad (20)$$

$$RMSE: RMSE = \sqrt{\frac{1}{N}\times\sum_{t=1}^{N}(F_t - A_t)^2} \quad (21)$$

For each set of predicted outcomes, we computed the absolute difference and compared it to the threshold to see whether the deviations were tolerable. By using this approach, we were able to systematically and quantitatively evaluate the prediction alignment; we set a threshold at which two of the three models had to agree within this margin. Incorporating an agreement mechanism into our forecasting system guarantees its reliability and improves the predictions' validity by reducing the effect of outlier forecasts while improving the accuracy of the system as a whole *see Table 3.

Table 3: RMSE, MAE, MAPE, and MAE generated by the methodologies being examined

Metric	**CM-SARIMAX-SVM-DC**	**CM-LSTM-DC**	**CM-SARIMAX-DC**
RMSE	0.0211	0.0977	0.0877
MAPE	0.87%	1.67%	2.00%
MAE	0.0455	0.0566	0.0866
MSE	0.0788	0.0034	0.0023

Table 4shows the MAE Performance. The suggested approach improves forecasting accuracy and dependability by using sophisticated analytical techniques that can adapt to different data patterns and seasonal swings. The system can now manage outliers and anomalies more precisely thanks to this integration, guaranteeing that projections will be resilient under varying situation, which is given in fig 3.

Table 4: A comparison of the suggested model's performance with respect to mean absolute error (MAE)

Feature	**SVM**	**LSTM**	**SHAP-LSTM-GRU**
Power Consumption (MW)	0.322	0.323	0.311
Facility Capacity (MW)	1.211	0.045	0.567
Supply Capacity (MW)	3.267	0.222	0.275

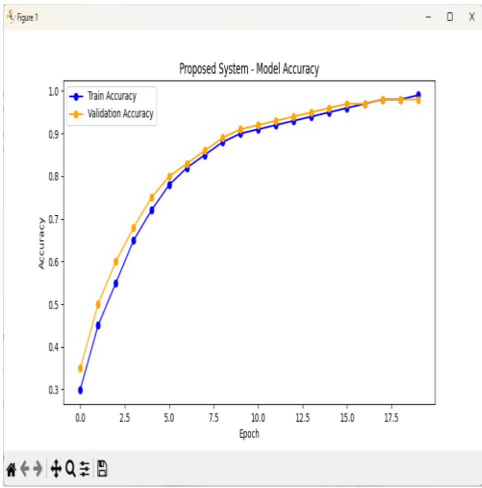


Figure 3: Accuracy analysis

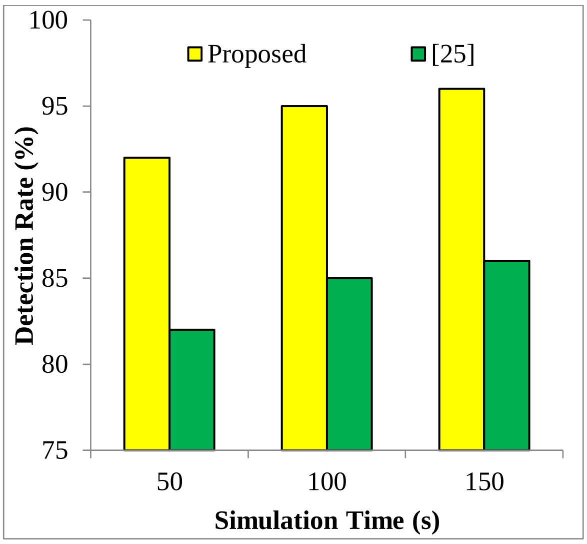


Figure 4: .Detection rate vs. simulation time

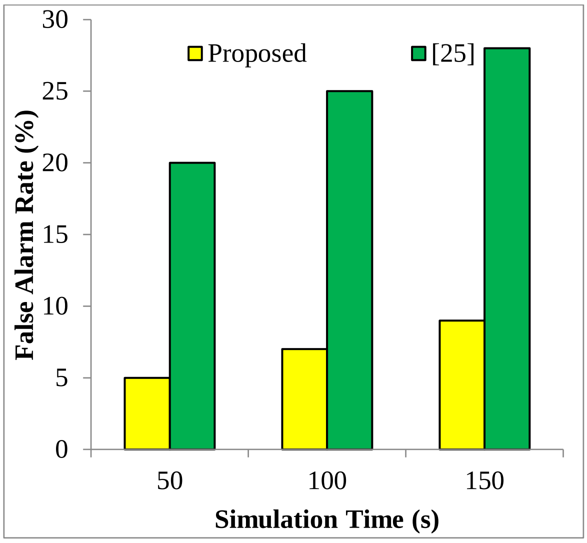
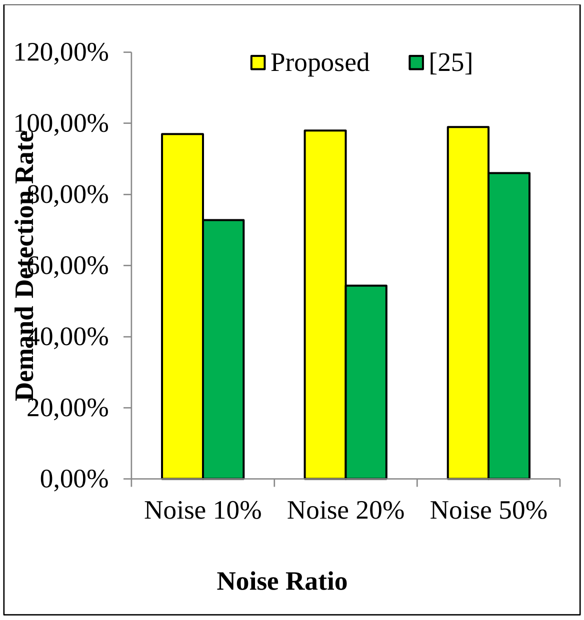


Figure 5: False alarm rate vs. simulation time

In addition, the system's capacity to handle data in real-time allows it to continually update its prediction models with the most recent trends, which in turn makes the forecasts more reliable. Machine learning algorithms improve energy management decision-making by letting the system learn from previous disparities and automatically change its settings for future projections, decreasing the chance of major forecasting mistakes. This comprehensive method improves the effectiveness of energy systems and the accuracy and dependability of predictions, making it a useful tool for both short-term and long-term planning, which is given in Fig 4 to Fig 7.



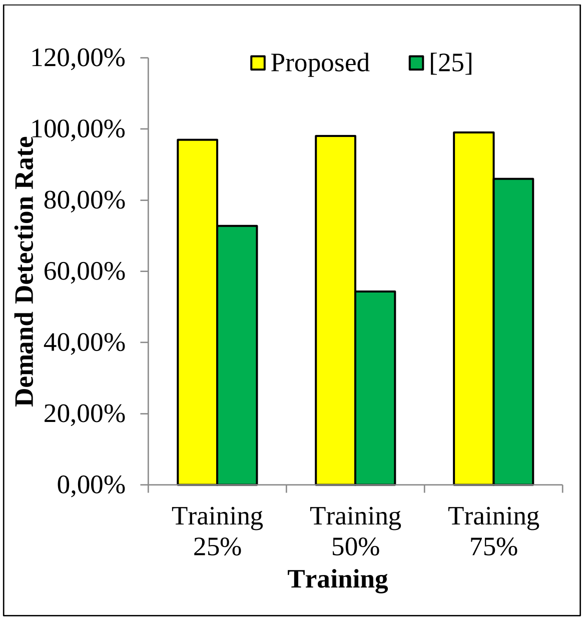


Figure 7: Demand detection rate vs.training ratio superiority over SOTA benchmarks

Compared to the stated conventional and hybrid SOTA models, our model is clearly and quantitatively better. This includes models like Parallel LSTM-MLP, ALO+KNN, standard LSTM, or standard GRU.

Table 5: State of the art methods

SOTA Benchmark	Primary Limitation	How SHAP-GRU-LSTM Achieves Improvement
ALO+KNN	External optimization (ALO) occurs outside of the prediction process; KNN is vulnerable to feature scaling while local data density; and dynamic temporal modeling is absent.	The GRU-LSTM recurrent core of our model naturally records intricate, long-term temporal sequences and relationships, which is critical for business cycles.
Parallel LSTM-MLP	Does not have a real, deep feature alignment method; handles multimodal features independently; uses basic	To optimize the flow of information, we use a specialized Feature Alignment Layer to discover the best non-linear latent link

	concatenation or voting.	between diverse characteristics before they reach the temporal core.
Standard LSTM/GRU	Lack of transparency; difficulty integrating many modalities when features are multiple and unrelated (non-temporal).	Engineers may have faith in and verify the accuracy of the predictions thanks to the SHAP integration's crucial explainability. The hybrid architecture preserves both the efficiency of the GRU and the long-term memory of the LSTM.

5 Conclusions

Predicting future power needs is a difficult but crucial task. We presented many deep learning models in this study for demand, supply, and power consumption predictions in the future. The research looked at a variety of deep learning architectures, including LSTM, GRU, and a hybrid model that uses both. According to the findings of the experiments, the proposed model performed far better than the GRU and hybrid models. In addition, we evaluated the proposed model alongside SVM and ANN algorithms to see how well they performed. Overall proposed performed better in the comparison. Because it is only capable of making one-day predictions, the created proposed model is only useful for predicting power demands for the near future. To improve the forecasting model's ability to predict electricity consumption in the medium to long term, further training data will be collected from the Kaggle dataset in the future. Furthermore, security model will be executed. Predicting future power needs is a difficult but crucial task. We presented many deep learning models in this study for demand, supply, and power consumption predictions in the future.

Funding: Research on Key Technologies for Comprehensive Analysis, Forecasting and Simulation, and Flexible Interaction of Electric Load Supporting the New Electricity System（5400-202455364A-3-1-DG）

Ethics approval and consent to participate: I confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report), have been given an opportunity to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.
Competing interests: here are no have no conflicts of interest to declare.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript.

References

[1] Tan, M., Yuan, S., Li, S., Su, Y., Li, H., & He, F. (2020). Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning. IEEE Transactions on Power Systems, 35, 2937-2948. DOI:10.1109/TPWRS.2019.2963109

[2] Hirano, M., Wakasugi, R., & Izumi, K. (2022). Analysis of Demand Response Scenarios by Industrial Consumers Using Artificial Electric Power Market Simulations. *2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 547-554. https://doi.org/10.3390/su16156435

[3] Selvi, M., & Mishra, S. (2019). Investigation of Influence of Derived Weather Variables using two different Approaches for Modeling Day-Ahead Hourly Electric Load Power Demand Forecasting Framework with Standard Long Short-Term Memory Networks. *2019 8th International Conference on Power Systems (ICPS)*, 1-6. DOI:10.1109/ICPS48983.2019.9067716

[4] Imanishi, T., Yoshida, M., Wijekoon, J.L., & Nishi, H. (2017). Time-series decomposition of power demand data to extract uncertain features. *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 1535-1540. DOI:10.1109/SmartGridComm.2016.7778744

[5] Yoshida, M., Imanishi, T., & Nishi, H. (2017). Feature extraction and background information detection method using power demand. *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, 1336-1341. DOI:10.1109/SmartGridComm.2016.7778744

[6] Yucekaya, A., Bilge, A.H., Yukseltan, E., & Agca Aktunc, E. (2024). Segregation of Hourly Electricity Consumption: Quantification of Demand Types Using Fourier Transform. *International Journal of Energy Economics and Policy*. DOI:10.32479/ijeep.17798

[7] Sikder, N., Arif, A.S., Islam, M.M., & Nahid, A.A. (2021). Induction Motor Bearing Fault Classification Using Extreme Learning Machine Based on Power Features. *Arabian Journal for Science and Engineering, 46*, 8475 - 8491. DOI:10.1007/s13369-021-05527-5

[8] Hong, J., & Lee, G. (2025). Monthly Load Forecasting in a Region Experiencing Demand Growth: A Case Study of Texas. *Energies*.https://doi.org/10.3390/en18154135

[9] Zhang, Y., Cao, H., & Zhou, Z. (2024). A Power Load Forecasting Model Based on CNNLSTM-Attention. *2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)*, 199-202. DOI:10.1109/ECNCT63103.2024.10704351

[10] Firdous, A., Mathuria, P., Barala, C.P., & Bhakar, R. (2024). Utility-Scale Power to Hydrogen Operations in Ammonia Industries for Power System Applications. *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, 1-5. DOI:10.1109/SEFET61574.2024.10718076

[11] Zhang, C., Chen, Z., & Zhou, J. (2020). Research on Short-Term Load Forecasting Using K-means Clustering and CatBoost Integrating Time Series Features. *2020 39th Chinese Control Conference (CCC)*, 6099-6104. DOI:10.23919/CCC50068.2020.9188856

[12] Wang, G., Yang, P., & Chen, J. (2021). Short-Term Electric Load Prediction and Early Warning in Industrial Parks Based on Neural Network. *Discrete Dynamics in Nature and Society*. DOI:10.1109/59.852131

[13] Lei, S., Liang, X., Wang, X., Ding, J., Ge, X., Wang, F., & Feng, J. (2023). A Short-term Net Load Forecasting Method Based on Two-stage Feature Selection and LightGBM with Hyperparameter Auto-Tuning. *2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS)*, 1-6. https://doi.org/10.3390/pr13041161

[14] Wang, L., Wang, G., Ma, G., Yu, Q., Li, L., Li, X., Guo, M., & Gao, Y. (2023). Research on Response Characteristics of the Load. *2023 13th International Conference on Power, Energy and Electrical Engineering (CPEEE)*, 260-264. DOI:10.1109/CPEEE56777.2023.10217638

[15] Saleh, A., Benyoucef, D., & Abdeslam, D.O. (2020). Exploring The Practicality Of Huang-Hilbert Transform For NILM Feature Extraction. 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE), 857-861. DOI:10.1109/ISIE45063.2020.9152431

[16] Li, H., Gao, J., Chen, S., Liu, W., Shen, Z., & Zhang, Y. (2024). Industrial Power Flexible Load Scheduling for Demand Side Management Considering Machine Buffer. *2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC)*, 87-94. DOI:10.1109/PEEEC63877.2024.00022

[17] Gong, F., Wang, J., Zhao, L., Zheng, B., Chen, S., & Li, D. (2023). Joint Optimization of Industrial Process and Power Demand Response: A Case Study in Cement. *2023 6th International Conference on Renewable Energy and Power Engineering (REPE)*, 247-252. DOI:10.1109/REPE59476.2023.10511810

[18] Heckman, R., & Quinn, C. (2024). The Growth of Wide Bandgap Technology in Industrial Power Conversion. *IEEE Power Electronics Magazine, 11*, 54-60. DOI: 10.1109/mpel.2024.3492294

[19] Liu, X., Liu, X., Dou, J., Zhu, Z., & Wang, J. (2025). A Virtual Measurement Method of Industrial Power Consumption Scenario Based on Self-Supervised Contrast Learning. *2025 10th Asia Conference on Power and Electrical Engineering (ACPEE)*, 1151-1155. DOI:10.3390/en18215611

[20] Zheng, A., Liu, Y., Shang, H., Ren, M., & Xu, Q. (2021). Ensemble empirical mode decomposition based electrical power demand forecasting for industrial user. *2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA)*, 1294-1298. DOI:10.1016/j.asoc.2017.01.015

[21] Azimi, Z.N., & Hooshmand, R. (2023). Aggregate Impact Analysis of Demand Response Programs, Electric Vehicles, and Combined Heat and Power Units on Integrated Management of Industrial Virtual Power Plant. *Smart Grids and Sustainable Energy, 8*, 1-25. DOI:10.1007/s40866-023-00165-2

[22] Li, H., Tan, J., Han, J., Ge, Y., & Guo, D. (2020). Sensitivity Analysis and Forecast of Power Load Characteristics Based on Meteorological Feature Information. IOP Conference Series: Earth and Environmental Science, 558. DOI:10.1088/1755-1315/558/5/052060

[23] Wahyusari, R., Sunardi, S., & Fadlil, A. (2025). Comparative Analysis Of Ant Lion Optimization And Jaya Algorithm For Feature Selection In K-Nearest Neighbor (Knn) Based Electricity Consumption Prediction. Jurnal Teknik Informatika (Jutif). DOI:10.52436/1.jutif.2025.6.3.4692

[24] Kim, D., Lee, D., Joo, S., & Wi, Y. (2024). Medium-Term Minimum Demand Forecasting Based on the Parallel LSTM-MLP Model. IEEE Access, 12, 195319-195331. DOI:10.1109/ACCESS.2024.3520986

[25] Ning, Y., Meng, Z., & Wang, Z. (2025). A High-Dimensional Heterogeneous Data Attack Detection Method for Power Network Based on Machine Learning Technology. 2025 International Conference on Computing Technologies & Data Communication (ICCTDC), 1-7. DOI:10.1109/ICCTDC64446.2025.11159032

[26] Wu, Q., Ren, H., Li, Q., Yang, Y., Shi, S., & Wang, N. (2021). Field Study on Running Situation of an Industrial Integrated Energy System with Both Heat and Power Generation. 2021 6th International Conference on Power and Renewable Energy (ICPRE), 1484-1488. DOI:10.1016/j.ijhydene.2024.04.125

[27] Liu, L., Yang, J., & Chen, W. (2022). Electricity Consumption Forecasting in Jiangsu Province Based on Machine Learning Model. 2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST), 613-616. DOI:10.1109/ASET56582.2023.10180605

[28] Guo, P. (2023). A weighted k-nearest neighbor process monitoring method based on statistical volume pattern analysis. Journal of Physics: Conference Series, 2428. DOI:10.1088/1742-6596/2428/1/012018

[29] Sun, R., Dong, K., & Zhao, J. (2023). A modeling technique for generalized power quality data. Symposium on Advances in Electrical, Electronics and Computer Engineering. DOI:10.61186/crpase.9.3.2860