

Data-Driven Load Prediction Models for Distribution Systems and Transformer Station Areas

Rui Du

ruidu10@outlook.com

Shanxi Datong University

Shangrong Liang

Shanxi Datong University

Junwei Tan

Shanxi Datong University

Yan Zhang

Shanxi Datong University

Guorui Liu

State Grid Hunan Electric Power Company Limited

Xin Wang

Shanxi Datong University

Yanfen Li

Shanxi Datong University

Research Article

Keywords: Deep Learning, Load Forecasting, Principal Component Analysis, Sliding Window, Smart Grid, Time Series Analysis

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10892170/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Data-Driven Load Prediction Models for Distribution Systems and Transformer Station Areas

Abstract

The forecasting accuracy of electrical load in distribution systems and transformer stations is essential to increase stability and reliability of the grid. Therefore, this research paper proposes a new hybrid forecast model based on temporal deep learning and feature refining techniques to forecast electrical load in distribution systems for a short-term period. In this case, data on historical electrical demand, meteorological variables and power network topology information are obtained from PJM Hourly Energy Consumption Dataset, MATPOWER IEEE Bus Dataset and NOAA Climate Dataset, correspondingly. For the purpose of improving data quality and stability before the learning process, preliminary preprocessing is conducted through interpolation, moving average, Savitzky-Golay filters, Zscore outlier detection, hourly time synchronization and Min-Max normalization. Afterwards, operational characteristics which consist of temporal, weather and electric components are extracted. Selection of features is conducted by means of Correlation Analysis (CA). Redundant variables are identified using Principal Component Analysis (PCA). Further, Sliding Window (SW) technique is employed to construct temporal sequence. In the forecasting architecture, there are two components: hybrid LSTM-XGBoost model, Long Short-Term Memory (LSTM) model used for learning temporal dependency and XGBoost regression algorithm for nonlinear load prediction. Also, hyperparameter tuning through Harmony Search Algorithm (HSA) is applied for improving the convergence rate and prediction accuracy. From experimental validation, it is evident that the proposed model outperforms in terms of accuracy and efficiency in predicting the model occurrence as compared to RF and SVM models with a MAE of 0.0210, MSE of 0.0006, MAPE of 0.0507, and R^2 of 0.9623.

Keywords: Deep Learning, Load Forecasting, Principal Component Analysis, Sliding Window, Smart Grid, Time Series Analysis

1. Introduction

The demand for electrical power is growing fast in modern electrical power networks, which are growing fast, because of industrial expansion, urbanization, population growth and increased use

of electrically powered technologies. Distribution systems and transformer station locations are vital for providing safe and reliable electrical service to the residential, commercial and industrial user (Pentsos et al., 2025). The distribution system is the last link in the power supply chain that includes feeders, substations, transformers, protection gear, and power lines to provide uninterrupted power to customers (Fang et al., 2024). The transformer station areas are significant operation units of the distribution networks since they control the voltage of the transmission line, provide electrical power to various service areas and stabilize the system when load variations occur (Zhuang et al., 2023), (Zhang et al., 2021). Environmental factors like temperature fluctuations, seasonal changes, weather events and natural disasters, have a profound impact on load demand and system performance, which in turn can impact distribution systems (Liu et al., 2026), (Huang et al., 2025). For energy planning, infrastructure extension, maintenance scheduling and demand side management, accurate knowledge of consumption pattern around the transformer station is required (Syed et al., 2021), (Özen et al., 2025). Operators of utilities and power systems are continuously gathering operational data at large scale from smart meters, sensors and supervisory control systems to gain better situational awareness and better decision-making process (Gong et al., 2022). Smart grid technologies have enabled the modernization of the power distribution infrastructure to allow for the transformation to be integrated with real-time communication, automated monitoring and intelligent control in the areas of transformer stations (Wang et al., 2023), (Tian et al., 2025). Smart distribution systems enable efficient fault identification, outage management, voltage optimization and energy utilization in interconnected distribution networks (Brahmachary and Ahmed, 2025), (He et al., 2022). Furthermore, the transformer stations are important nodes which affect the overall stability and reliability of the grid during peak load situations (Fan et al., 2024), (Zabin et al., 2024). Efficient operation of distribution systems can help to minimize transmission losses, optimize energy use, and help to ensure customer satisfaction (He et al., 2023), (Gao et al., 2025). The use of digital technologies and smart monitoring has changed the traditional distribution networks into data-driven energy management spaces (Zhao et al., 2023). However, the growing demand for electricity and variable load complexities still poses operational and planning challenges to utilities. Enhancing the reliability, stability and operation efficiency of distribution systems and transformer station areas is therefore still a great research and industrial challenge in the field of power engineering in modern applications (Zhang et al., 2024), (Zhao et al., 2021).

1.1 Problem statement

Many load forecasting (LF) models for distribution systems and transformer station areas have various shortcomings in accurately modeling complex temporal load variations, nonlinear load consumption behavior, and dynamic environmental load effects. In the case of high dimensional sequential load data that has seasonal patterns and rapidly changing operating conditions, traditional statistical and machine learning methods (e.g., RF and SVM) may have low forecasting performance. A number of current algorithms also consider only a small number of representations and are unable to comprehensively leverage the information on weather, transformer operational features, and the topology of the power network in the forecasting process. Moreover, insufficient pre-processing, sub-optimal feature selection and sub-optimal hyperparameter tuning makes the computations complex and the prediction stability is not ensured (Zhang et al., 2024), (Zhao et al., 2021).

1.2 Novelty

The novelty of the proposed work is that a unified intelligent forecasting framework that combines CA-based feature selection, PCA-based dimensionality reduction, SW-based temporal sequence construction, hybrid LSTM–XGB forecasting and HSA-based hyperparameter optimization within a single architecture for distribution system and transformer station load prediction has been developed. Compared to traditional forecasting models, the proposed framework not only uses historical demand information but also considers temporal behavior, weather conditions, electrical operational characteristics, etc. to enhance the reliability of forecasting and the ability of the model to learn non-linear relationships. Furthermore, load mapping and load overloading of transformer stations are also included in the forecasting process, which allows for intelligent analysis of the operation in the transformer station and practical smart grid energy management with higher forecasting accuracy, convergence stability and computational efficiency.

1.3 Contribution

- Developed a hybrid LSTM-XGB forecasting framework which can predict load accurately in the distribution system and transformer station areas.

- Combined feature selection using integrated CA and dimensionality reduction using PCA to remove attributes with no contribution to the forecasting performance and solve the curse of dimensionality.
- Developed SW based temporal learning sequences to learn patterns of historical demands and seasonal consumption patterns.
- Used HSA technique to optimize hyper-parameters which improved convergence stability and reduced the forecasting error.
- Identify loading operating conditions of a transformer station for overload detection, peak demand monitoring and operating assessment.

1.4 Structure of the paper

The structure of the remaining part of this paper will be as follows: Section 1 includes background, problem statement, novelty and contribution, Section 2 includes literature survey, Section 3 includes proposed methodology, Section 4 includes experiment results and performance evaluation, and Section 5 includes conclusion and limitations of the research with future scope.

2. Literature survey

Recent studies are discussed in this section, covering different machine learning and deep learning methods to Load Prediction Models for Distribution Systems and Transformer Station Areas, their pros and cons.

In (Ramesh et al., 2022) presented a machine learning based approach for load monitoring and management in smart distribution transformers. The goal was to study and avoid overloading conditions for transformers. The method enhanced the efficiency of monitoring and reliability of operations through the intelligent prediction techniques. But the model primarily was used to monitor tasks and was not very capable of long-term forecasting when the load was changed dynamically. Rung-Fang (Chang et al., 2002) proposed a load modeling method for distribution transformer based on load research data after studying the transformers. The goal was to analyze the loading characteristics of the transformers according to customer demand. The model showed better load representation accuracy and facilitated the use of distribution planning applications. But the technique was very dependent on historical data and was not flexible enough in adapting to the fast-evolving load conditions of the smart-grid. In order to improve the traditional load

modeling methods, Alexandra (L'Heureux et al., 2022) proposed a transformer-based deep learning model for electrical load forecasting. The goal was to learn about long-range temporal dependency with the help of self-attention. The model has been found to be more accurate and more scalable than the recurrent models. However, during practical implementation, a high number of computational resources were needed and training complexity was added to the transformer architecture. Building on this research, (Grabner et al., 2023) introduced a single framework for distribution network load forecasting. The goal was to get a model that could be generalized for multiple transformer regions. The framework enhanced the scalability and power of forecasting in smart-grid systems. The generic structure, however, may not account for the local load characteristics, and may also lead to inaccurate predictions in non-homogeneous operation, H. (Gorginpour et al., 2022) introduced a single framework for distribution network load forecasting. The goal was to get a model that could be generalized for multiple transformer regions. The framework enhanced the scalability and power of forecasting in smart-grid systems. The generic structure, however, may not account for the local load characteristics, and may also lead to inaccurate predictions in non-homogeneous operation. Similarly, (Torres-Bermeo et al., 2025) proposed clustering and machine learning models for transformer load curve characterization. The objective was to identify representative load patterns and improve transformer sizing accuracy. The framework enhanced load analysis and transformer planning efficiency. However, clustering performance strongly depended on dataset quality and parameter selection strategies. Further improving forecasting reliability, (Bouhamed et al., 2025) to enhance the reliability of forecasting. The goal was to enhance the uncertainty aware forecasting in presence of dynamic demand variations. The framework improved the reliability of prediction and stability of probabilistic forecasting. Additionally, the combination of the transformer and probabilistic learning mechanisms resulted in high computational complexity and training costs. In addition to forecasting applications, A.S. (Mogos et al., 2023) suggested a hybrid one-class deep SVDD (SSVDD) model for transformer failure prediction and predictive maintenance. The goal was to use a set of lightning strike failure data to identify abnormal operating conditions. The framework increased the reliability of maintenance and minimized unplanned transformer outages. The model, however, is found to have limited generalization ability for other fault conditions in the transformer and is mainly limited to failures caused by lightning. In order to achieve long-term forecasting improvement, (Zhan et al., 2022) introduced a hierarchical decomposition self-attention network

for energy load trend prediction. The objective was to capture global and local temporal patterns for reliable smart-grid forecasting. The framework improved long-term forecasting precision and nonlinear load modeling capability. However, the architecture increased parameter optimization difficulty and computational burden during deployment. Recently, A. (Hussain et al., 2025) introduced an LSTM model coupled with the transformer as a hybrid forecasting model for electric vehicle charging demand. The goal is to predict the charging demand with temporal learning and attention mechanisms. Comparative Study of Existing Models for Load Forecasting and Transformer Management is illustrated in Table 1.

Table 1: Comparative Analysis of Existing load Forecasting and Distribution Transformer Management Models

Author	Method	Objective	Advantage	Limitation
(Ramesh et al., 2022)	Machine Learning Framework	To monitor transformer load and prevent overloading conditions	Improved monitoring efficiency and operational reliability	Limited long-term forecasting capability under dynamic load variations
(Chang et al., 2002)	Load Modeling Technique	To estimate transformer loading characteristics using historical data	Enhanced load representation and planning accuracy	Strong dependency on historical datasets
(L’Heureux et al., 2022)	Transformer-Based Deep Learning	For capturing long-term dependencies in load forecasting	Improved forecasting accuracy and scalability	High computational complexity and training cost
(Grabner et al., 2023)	Global Forecasting Framework	To develop a unified forecasting model	Improved scalability and forecasting robustness	Reduced effectiveness for localized load behavior

		for multiple regions		
(Gorginpour et al., 2022)	Intelligent Maintenance Scheduling Model	To estimate transformer lifetime and optimize maintenance scheduling	Enhanced transformer reliability and sustainability	Requires highly accurate environmental data
(Torres-Bermeo et al., 2025)	Clustering with Predictive ML Models	To characterize load curves and improve transformer sizing	Improved transformer planning efficiency	Performance depends on dataset quality
(Bouhamed et al., 2025)	Deep Probabilistic Transformer Network	To perform uncertainty-aware load forecasting	Enhanced prediction reliability and probabilistic stability	Increased computational and training complexity
(Mogos et al., 2023)	Hybrid One-Class Deep SVDD	To predict transformer failures for predictive maintenance	Improved abnormality detection and maintenance reliability	Limited generalization for different fault conditions
(Zhan et al., 2022)	Hierarchical Self-Attention Network	To improve long-term smart-grid load forecasting	Enhanced long-term forecasting precision	High parameter optimization complexity
(Hussain et al., 2025)	LSTM-Based Hybrid Transformer	To forecast EV charging station demand	Captured long-range dependencies effectively	Requires large-scale datasets and high computation

Transformer networks, clustering, and self-attention mechanisms have demonstrated their effectiveness in transformer load monitoring, load forecasting, maintenance scheduling and predictive analytics using machine learning, but there are still several research limitations. The current models are primarily based on monitoring or forecasting or maintenance without providing an integrated intelligent framework for comprehensive load prediction and management of the transformer. Further, a number of the deep learning and transformer-based approaches demand heavy computing power and massive amounts of data, which hinders real-time applications in real-world distribution systems. The current methods also exhibit lower adaptability when the load changes, renewables are added, environmental conditions change, and the behaviors of the transformers are not uniform. Moreover, these are the characteristics of the load not well explored and temporal dependency learning, feature redundancy reduction and uncertainty-aware prediction have not received enough attention, which motivates the development of an efficient, scalable and robust hybrid framework for forecasting of smart distribution transformer applications.

3. Proposed Methodology

The suggested approach is to build an intelligent data-driven load forecasting system for distribution grid and transformer station area based on the temporal, environmental and electrical operational data. Historical load demand is obtained from the PJM dataset, while the NOAA dataset provides weather data and the MATPOWER IEEE 33-bus dataset provides electrical operational parameters. The incorporation of MATPOWER simulation information into PJM and NOAA data allows for the framework to learn the temporal demand behavior, environmental influences and power system operational characteristics together for better transformer station load forecasting and overload analysis. The operational features are extracted at the temporal, weather and electrical domains and then feature selection is done with CA and dimensionality reduction with PCA. The SW techniques are employed to create sequential load patterns and a hybrid forecasting framework based on LSTM and XGB is developed for temporal learning and nonlinear load prediction. Finally, the hyperparameter optimization based on HAS enhances forecasting accuracy, the convergence performance and forecasting reliability. The proposed model is illustrated in the block diagram of Figure 1.

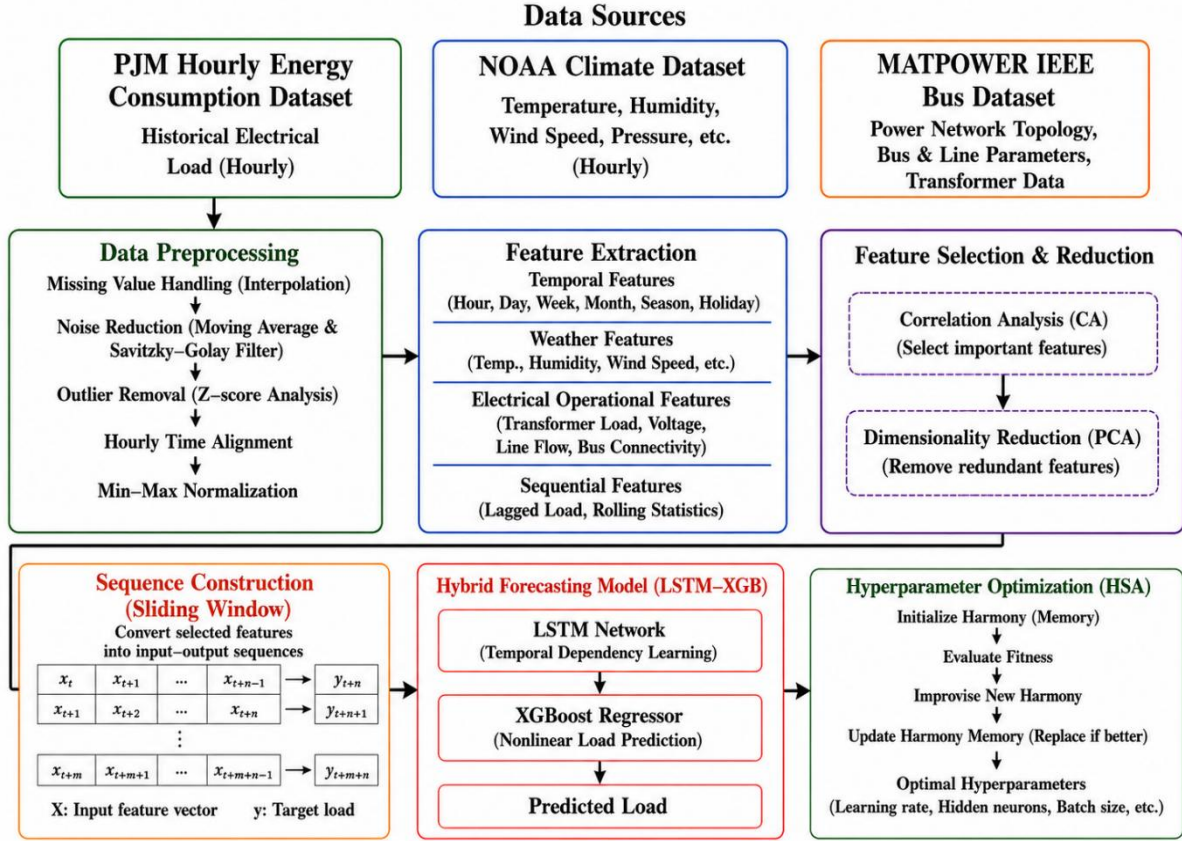


Figure 1: Block diagram of proposed model

3.1 Data Acquisition

During the data acquisition stage, the historical electrical demand, weather information, and operational information of the power network are collected from open data sources.

3.1.1 PJM Hourly Energy Consumption Dataset

The PJM Hourly Energy Consumption Dataset (Zhan et al., 2022) is collected from data points throughout the world from stations operated by NOAA that monitor weather from various countries and climatic zones. The data base consists of several million samples of meteorological observations from international climate stations over many decades. The major features are: Temperature (°C), Atmospheric Pressure, Dew Point, Wind Speed (km/h) Visibility, Precipitation, Station ID and Humidity (%).

3.1.2 NOAA Climate Dataset

The NOAA Climate Dataset (Hussain et al., 2025) is collected from the National Oceanic and Atmospheric Administration global weather monitoring stations distributed across different countries and climatic regions. The dataset contains several million meteorological observation samples recorded over multiple decades from international climate stations.

3.1.3 MATPOWER IEEE Bus Dataset

The MATPOWER IEEE Bus Dataset available from the MATPOWER simulation platform, includes several IEEE bus systems, including the IEEE 9-bus, 14-bus, 33-bus, 57-bus and 118-bus systems, all having hundreds of electrical operational samples. Here, IEEE bus 33 system is used. Some of the important features are: Bus Voltage Magnitude (p.u), Active Power Demand (MW), Reactive Power Demand (MVAR), Voltage Angle (Degree), Generator Output, Branch Resistance, Transformer Connections, Transmission Line Parameters. Actual values are a voltage magnitude of 1.02 p.u., an active power demand of 90MW, and a reactive power demand of 30MVar.

3.2 Advanced load Preprocessing Methodology

In this section, Data Preprocessing Workflow, the preprocessing workflow is presented as a sequence of transformations on the raw time-series load data to generate a clean, temporally aligned, noise reduced and normalized representation, suitable for stable learning. In the first stage, missing value handling is done by interpolating missing observations in the raw load sequence. The reconstructed signal is obtained in equation (1).

$$\hat{x}_t = m_t x_t + (1 - m_t) \tilde{x}_t, \quad (1)$$

where the interpolated estimate is $\tilde{x}_t$. Here m_t and x_t represent nearest valid observations. The second stage uses moving average smoothing to reduce the high-frequency noise and retain overall trend information. The smoothed signal is given by the equation (2).

$$s_t = \frac{1}{k} \sum_{i=0}^{k-1} x_{t-i}, \quad (2)$$

where k is the size of the sliding window used to smooth and eliminate stochastic noise components in the load variation patterns. The third stage uses Savitzky–Golay filtering for polynomial regression based smoothing. The filtered output is shown in equation (3).

$$y_t = \sum_{i=-n}^n c_i x_{t+i}, \quad (3)$$

where c_i are the coefficients of the convolution which are obtained by the least-squares polynomial fitting on a window of length $2n + 1$. Z-score based outlier detection and correction are used to ensure the statistical consistency. The standardized score is computed in equation (4).

$$z_t = \frac{x_t - \mu}{\sigma}, \quad (4)$$

where μ and σ are mean and standard deviation of the signal. The fifth stage is time alignment by hour, which involves evenly sampling irregular observations. The aligned signal is given by equation (5).

$$x_h = \frac{1}{|S_h|} \sum_{t \in S_h} x_t, \quad (5)$$

This is to ensure that the temporal resolution is uniform across all the samples S_h within hour h , necessary for sequential modeling. The last layer normalizes all features between $[0,1]$ using Min–Max normalization, which makes the numerical values more stable. The transformation is given by equation (6).

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6)$$

Where, x_{min} and x_{max} are the minimum and maximum values for the features, which is important for getting uniform scaling in all the input dimensions. Lastly, the embedded preprocessing scheme is able to handle missing data, noise, outliers and temporal inconsistencies of the load data, which guarantees the reliability of the data and contributes to more stable learning.

3.3 Advanced Feature Extraction Methodology

In this section, Feature Extraction Process, a detailed feature engineering process is developed to extract highly discriminative temporal, meteorological, and electrical operational features from the preprocessed load dataset Temporal feature extraction is first used to encode periodic load dynamics related to the hourly, daily, and seasonal consumption characteristics. Sinusoidal temporal encoding is formulated in (7) in order to maintain the continuity of the cycles and to prevent the discontinuity of the temporal intervals at the terminals.

$$h_t^{sin} = \sin\left(\frac{2\pi h_t}{24}\right), h_t^{cos} = \cos\left(\frac{2\pi h_t}{24}\right) \quad (7)$$

where, h_t denotes the hourly index corresponding to daily operational behavior. Similarly, weekly consumption periodicity is represented using equation (8).

$$d_t^{sin} = \sin\left(\frac{2\pi d_t}{7}\right), d_t^{cos} = \cos\left(\frac{2\pi d_t}{7}\right) \quad (8)$$

where d_t represents the day index associated with weekly demand variability. The integrated meteorological influence is represented using equation (9).

$$W_t = \sqrt{T_t^2 + H_t^2 + R_t^2} \quad (9)$$

where, W_t is the aggregated climatic impact factor, which is the sum of the atmospheric effect on energy demand. Moreover, the modeling of the nonlinear interaction between the environmental conditions and the load consumption is provided by the equation (10).

$$F_t = x_t(T_t + H_t) \quad (10)$$

where, F_t is the weather-load interaction feature that indicates the weather-dependency of electricity consumption. The apparent power magnitude is computed using equation (11).

$$S_t = \sqrt{P_t^2 + Q_t^2} \quad (11)$$

where, S_t represents the total apparent power associated with transformer operating conditions. Transformer utilization characteristics are further estimated using equation (12).

$$U_t = \frac{S_t}{S_{rated}} \quad (12)$$

Where, S_{rated} denotes the rated transformer capacity and U_t represents transformer utilization ratio reflecting operational stress conditions. Further, the variability in loading is calculated in a sequential way by means of equation (13).

$$\Delta x_t = x_t - x_{t-1} \quad (13)$$

where Δx_t is the instantaneous load change between successive temporal intervals. Finally, the temporal, environmental and operational features are extracted and combined into a multidimensional matrix of features to improve the discriminative learning capability.

3.4 Correlation and PCA Optimization

In the section, CA and PCA are used to find the features that are most informative for forecasting, to remove redundant and noise-sensitive features, and to create a compact feature representation.

3.4.1 Correlation-Based Feature Selection

The first step is to use the correlation analysis to assess statistical dependency between extracted features and the forecasting target variable. The Pearson correlation coefficient is computed using equation (14).

$$\rho_{f_i, Y} = \frac{\sum_{t=1}^T (f_{it} - \bar{f}_i)(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^T (f_{it} - \bar{f}_i)^2} \sqrt{\sum_{t=1}^T (Y_t - \bar{Y})^2}} \quad (14)$$

where $\rho_{f_i, Y}$ denotes the correlation coefficient between the i^{th} feature and forecasting target, f_{it} represents the feature value at time t , and $\bar{f}_i$ and $\bar{Y}$ are the average of the feature and the forecasting target, respectively. The attributes having larger absolute correlation coefficients are kept as dominant descriptors to forecast, while the attributes having low correlation and redundancy are discarded.

3.4.2 Principal Component Dimensionality Reduction

Then, PCA is used to create an orthogonal representation of the optimized feature subset with maximal variance information. Suppose the feature matrix is centralized and they denoted as F_c . Covariance matrix that describes the dependency between features is calculated using equation (15).

$$C = \frac{1}{m-1} F_c^T F_c \quad (15)$$

where, C is the covariance matrix and m denotes the number of observations. Eigenvalue decomposition is then carried out to extract the directions of dominant variance from (16).

$$C v_i = \lambda_i v_i \quad (16)$$

where, λ_i and v_i denote the eigenvalue and corresponding eigenvector of the i^{th} principal component. Then the obtained reduced feature representation is subbed into equation (17).

$$Z = F_c V_k \quad (17)$$

The projection matrix is denoted as V_k represented by the principal eigenvectors selected and the low-dimensional feature matrix is denoted as Z denotes the feature matrix after the transformation, which retains the main matrix of the statistical information, so it can be used for efficient forecasting analysis. Redundant and highly correlated features are removed and key structural information is retained through PCA-based transformation.

3.5 Temporal Sequence Construction Process

This section is called Temporal Sequence Construction Process, and uses the sliding window (SW) mechanism to convert the extracted multidimensional feature representation into a series of temporal samples for supervised forecasting analysis. The input sequence generation process is formulated in equation (18).

$$X_t = [f_{t-w+1}, f_{t-w+2}, \dots, f_t] \quad (18)$$

where $X_t \in R^{w \times n}$ denote the extracted feature sequence, with f_t being the n dimensional feature vector at time instant t . The process of generating input sequence is written as follows: equation (19).

$$Y_t = f_{t+\Delta} \quad (19)$$

where, Y_t represents the prediction target corresponding to forecasting horizon Δ , while $f_{t+\Delta}$ to be predicted by the predictive model. From this point on, all the temporal samples produced are summed into a supervised learning set (formulated as equation (20))

$$D = \{(X_t, Y_t)\}_{t=w}^{T-\Delta} \quad (20)$$

where, D denotes the final sequential training dataset, T represents the total number of temporal observations, and each ordered pair (Y_t) are the forecasting sample. The sliding window formulation systematically preserves temporal continuity, sequential dependency characteristics.

3.6 Hybrid LSTM–XGB Forecasting Framework

In this section, Hybrid LSTM–XGB Forecasting Framework, a hybrid predictive architecture is proposed by combining LSTM networks with XGBoost regression to model both the long-range temporal dependencies as well as the characteristics of load variation in the energy consumption sequences.

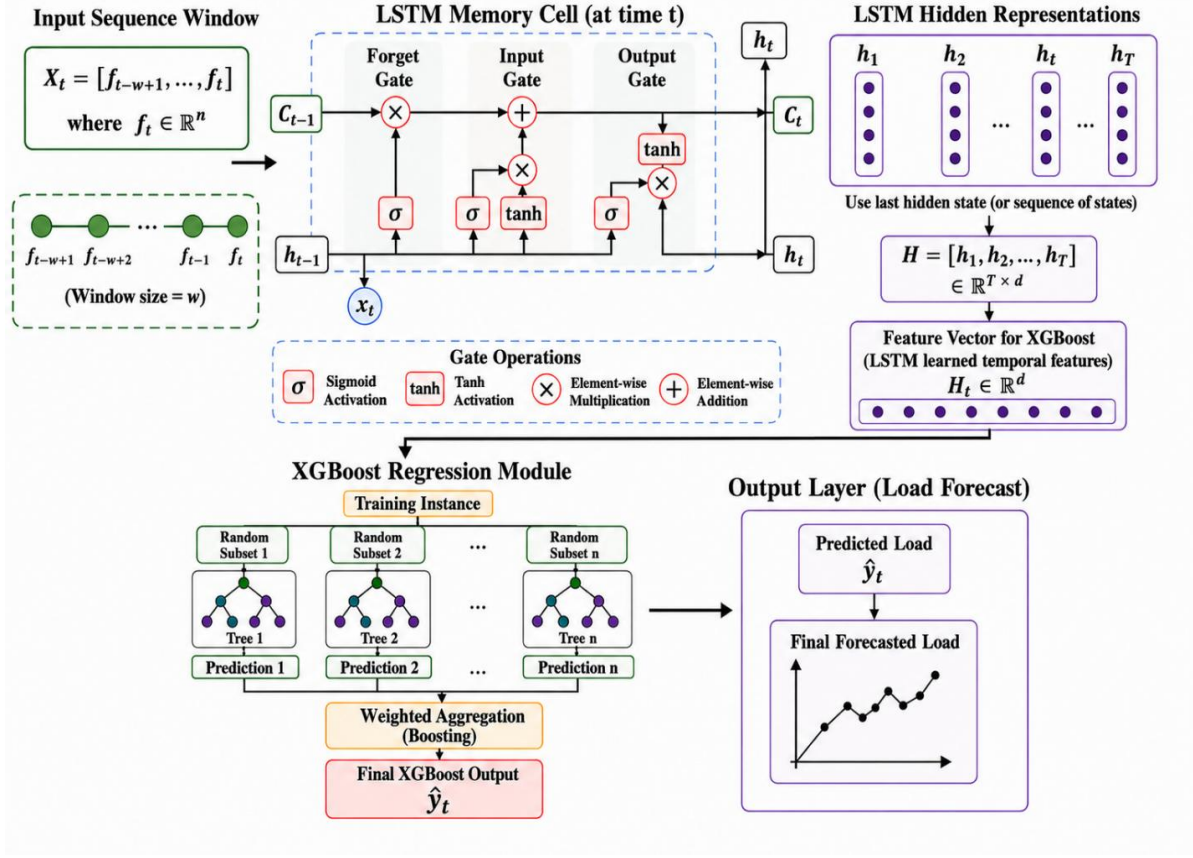


Figure 2: Architecture of the proposed hybrid LSTM–XGBoost framework for load forecasting

The proposed hybrid model architecture is designed to carry out load forecasting sequentially through temporal feature extraction and ensemble-based regression as shown in figure 2. To address the long-term temporal dependence, stacked LSTM layers are employed to process the historical load sequences and then the hidden representations learned from the LSTM layers are converted to discriminative feature vectors input to an XGBoost regression framework, which enables robust nonlinear forecasting with superior predictive generalization and accuracy. The memory state propagation scheme of LSTM architecture is controlled by equation(21).

$$h_t = o_t \odot \tanh(C_t) \quad (21)$$

where, h_t represents the hidden state vector, o_t denotes the output gate activation, C_t corresponds to the cell memory state, and $\odot$ indicates element-wise multiplication. The memory cell state is recursively updated using equation (22).

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (22)$$

Where, f_t and i_t denote forget and input gate activations, respectively, while $\tilde{C}_t$ represents candidate memory information extracted from the current temporal observation. Let $H = \{h_t\}_{t=1}^T$ denote the learned temporal feature sequence obtained from the recurrent architecture. The final prediction function of XGBoost is formulated using equation (23).

$$\hat{y}_i = \sum_{k=1}^K f_k(h_i) \quad (23)$$

Where, $\hat{y}_i$ represents the load demand predicted from input representation h_i , K is the number of boosted regression trees, and f_k is the non-linear decision function for the k^{th} tree learner. The formula (24) is used to represent the overall objective function optimized in the gradient boosting.

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (24)$$

where, y_i and $\hat{y}_i$ denote actual and predicted load values, $l(\cdot)$ represents the regression loss function, and $\Omega(f_k)$ is the regularization term that regulates the model complexity and overfitting behavior. Each regression tree, whose structure has been regularised, is further specified by equation (25).

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (25)$$

where T denotes the number of leaf nodes, w_j represents the weight associated with the j^{th} leaf, and γ and λ are regularization coefficients for the structural sparsity and weight penalization, respectively. Last but not least, the hybrid integration of LSTM-based temporal sequence modeling and XGBoost-based nonlinear regression allows the forecasting framework to effectively capture the complex sequential dependencies, transformer load variation under dynamically varying operational conditions.

3.7 Harmony Search Based Hyperparameter Optimization

In this part, Harmony Search Optimization Process, the HSA is used to optimize the selection of some critical hyperparameters such as learning rate, number and configuration of hidden neurons and batch size to stabilize the convergence of the algorithm, reduce forecasting error and improve the generalization performance.

Step 1: Harmony Memory Initialization

The hyperparameter vector is denoted as $H = [\eta, n_h, b_s]$, where η denotes learning rate, n_h represents the number of hidden neurons, and b_s corresponds to batch size. Harmony memory matrix with the good candidate solution vectors is generated by equation (26).

$$HM = [H_1 H_2 : H_m] \quad (26)$$

where HM denotes harmony memory and m represents harmony memory size. To ensure that the parameters explored during the optimization are feasible, each harmony vector is generated in the preset lower and upper search space.

Step 2: Random Harmony Generation

Harmony Memory consideration mechanism and random exploration mechanism are performed to obtain a new candidate solution to increase the search diversity after harmony memory initialization. The harmonics vector, which is newly generated, is represented in equation (27).

$$H_{new} = H_r + \alpha(2r - 1) \quad (27)$$

A randomly selected harmony vector from memory is denoted as H_r , pitch adjustment bandwidth is denoted as α , and r is uniformly distributed random variable in $[0,1]$. Figure 3 shows Flow chart of Harmony Search Based Hyperparameter Optimization

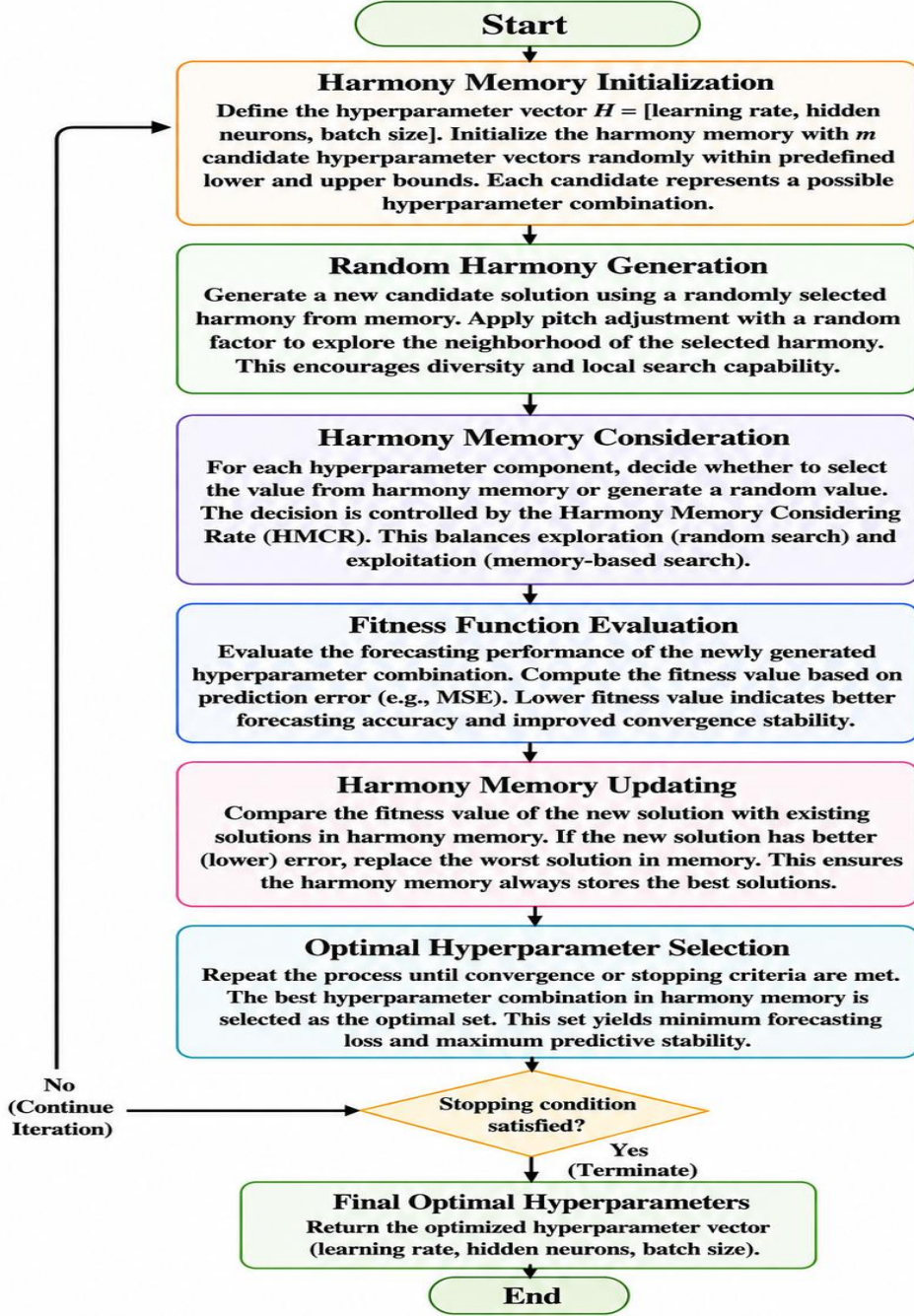


Figure 3: Flow chart of Harmony Search Based Hyperparameter Optimization

Step 3: Harmony Memory Consideration

Then harmony memory consideration is done to decide whether to choose the new hyperparameter values from the harmony memory or just random values. The selection mechanism is designed by solving equation (28).

$$P(H_i) = \{HMCR, memory\ selection\ 1 - HMCR, random\ selection\} \quad (28)$$

where, $HMCR$ denotes Harmony Memory Considering Rate controlling the probability of selecting existing candidate solutions from memory.

Step 4: Fitness Function Evaluation

After generating the harmonies, a fitness function minimising the prediction error is used to test the forecasting performance of each hyperparameter combination. The objective function is evaluated with the formula in (29).

$$J(H) = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (29)$$

where, $J(H)$ denotes optimization fitness value, y_i represents actual load demand, $\hat{y}_i$ denotes predicted output, and N indicates the number of training samples.

Step 5: Harmony Memory Updating

In order to update the harmony memory, the inferior candidate solutions are replaced with newly generated candidate solutions with smaller forecasting errors after fitness evaluation. The equation (30) depicts the optimum harmony memory update process.

$$HM^* = \arg \arg J(H_i)_{H_i} \quad (30)$$

where HM^* denotes the updated optimal harmony memory and H_i represents individual candidate hyperparameter combinations.

Step 6: Optimal Hyperparameter Selection

Finally, the optimization process reaches a minimum of forecasting loss and maximum of predictive stability with the best hyperparameter configuration.

4. Result and Discussion

The model is implemented in python 3.11.9, TensorFlow 2.20.0, XGBoost 3.1.3, Scikit-learn 1.5.2, Pandas 3.0.0 and NumPy 2.3.5 libraries. The implementation is carried out on a desktop system with a 64-bit Windows 10 operating system. Efficient model training and a large-scale temporal data processing are achieved by the hardware configuration comprising of the Intel64 Family 6 Model 183 Genuine Intel processor featuring 24 physical cores and 32 processing threads and 31.72 GB RAM. Furthermore, same procedure and evaluation settings are always used for all

comparative models to guarantee the performance evaluation. The proposed model's Hyperparameter is given as Table 2.

Table 2: Hyperparameter of proposed model

Hyperparameter	Value
Hidden Units	64
Learning Rate	0.001
Dropout Rate	0.2
Sequence Length	24 Hours
Batch Size	32
Epochs	100
Training Split	80%
Testing Split	20%
Sliding Window Size	24

4.1 Performance measures

The proposed load forecasting model is tested with the standard classification metrics: RMSE, MAPE and MAE, R^2 which all evaluate the effectiveness and reliability of the model in load forecasting in transformer areas.

4.1.1 MSE

MSE is used to measure the average magnitude of prediction error between the actual load and the estimated loads generated by the proposed model. The RMSE is calculated in equation (31).

$$MSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (31)$$

where , N represents the total number of samples, y_i indicates the actual observed state value, and $\hat{y}_i$ represents the predicted value generated by the model.

4.1.2 MAE

MAE measures the average absolute difference between the actual values and the predicted values obtained from the load forecasting framework. It is mathematically expressed in equation (32).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (32)$$

4.1.3 MAPE

MAPE is used to measure the average relative difference between actual and predicted values and it is computed in equation (33).

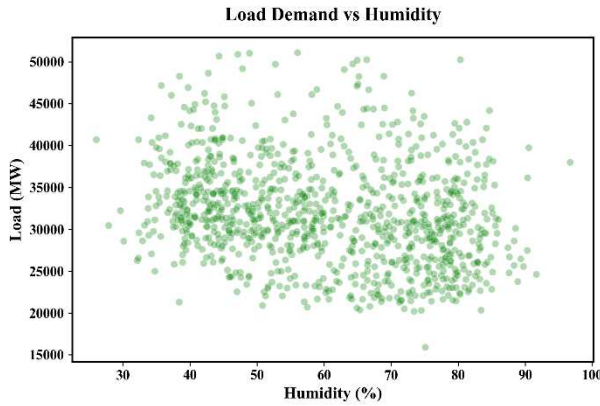
$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (33)$$

4.1.4 R^2

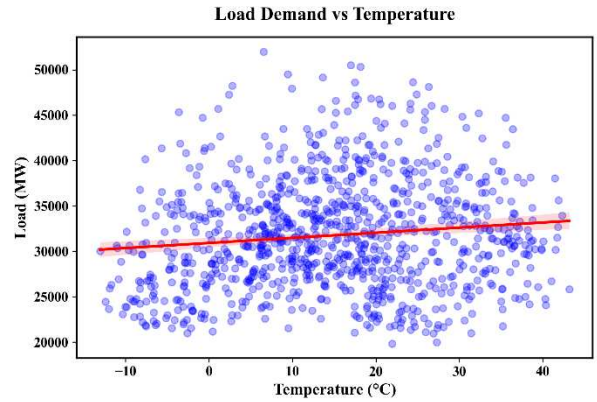
R^2 used to measure the strength of the relationship between predicted values and actual values. It is represented in equation (34).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (34)$$

4.2 Experimental results of proposed model



(a)



(b)

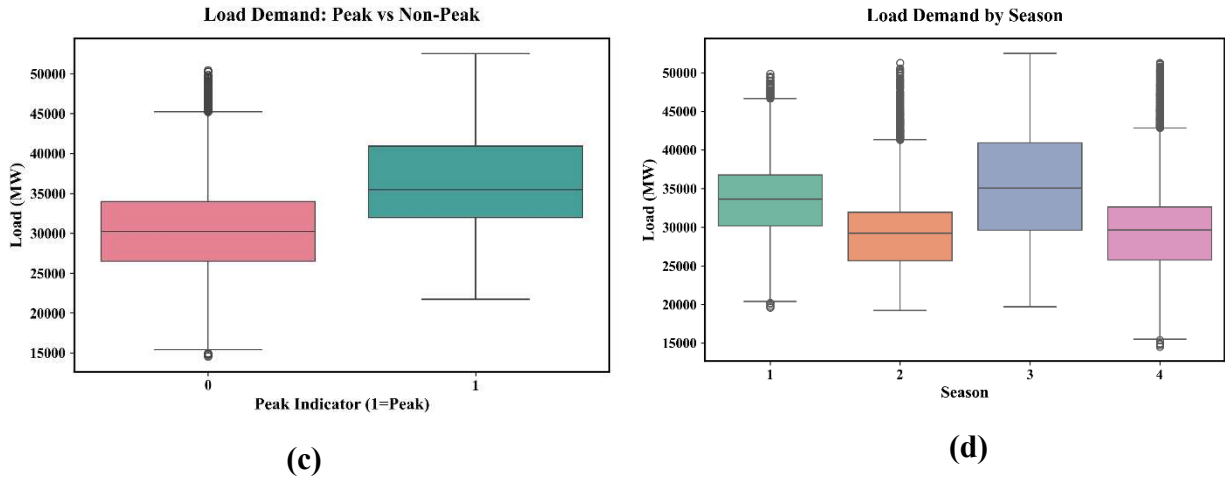
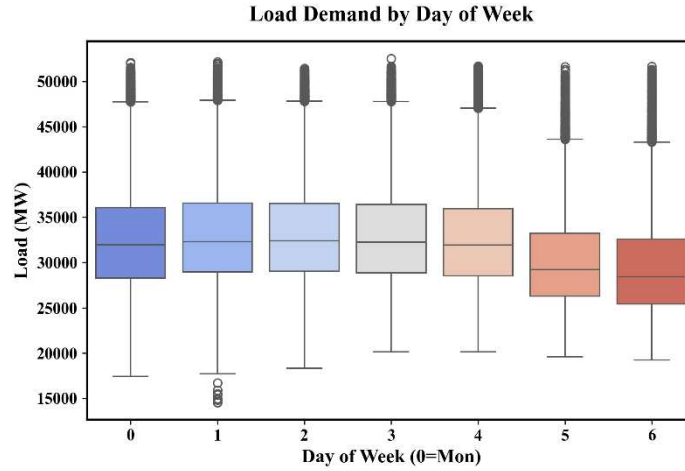
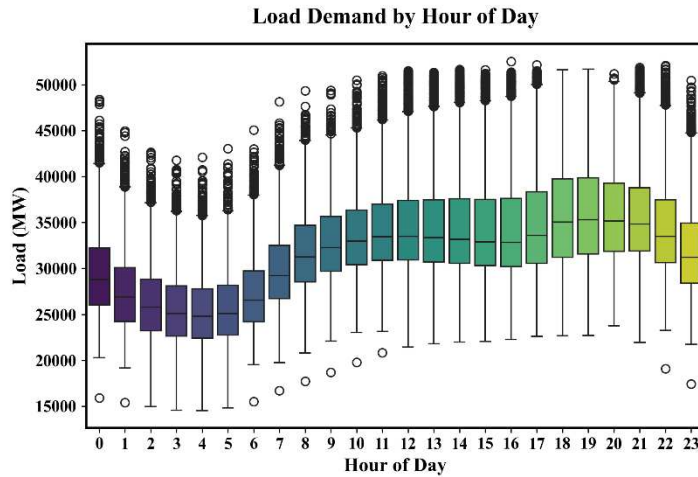


Figure 4: Exploratory Data Analysis of (a) Load Demand vs. Humidity, (b) Load Demand vs. Temperature (c) Load Demand: Peak vs. Non-Peak (d) Load Demand by Season

Figure 4 shows an exploratory data analysis of Load Demand vs. Humidity that indicates that there is not a strong relationship between humidity and load demand, as the load values are scattered broadly from around 15,000 MW to 51,000 MW for a range of humidity between 30% and 100%. The Load Demand vs. Temperature scatter plot, on the other hand, shows a slight upward linear trend, with a red regression line showing an increase in electricity demand from approximately 30,000 MW at -10°C to approximately 33,000 MW at 40°C , suggesting that a positive relationship between temperature and electricity demand exists, probably owing to cooling loads. It is clear from the box plot of Load Demand: Peak vs. Non-Peak that there is a significant difference between the loads registered during peak hours (indicator = 1), with a median load around 35,500 MW and IQR from 32,000 MW to 40,500 MW, and the loads registered during non-peak hours (indicator = 0), with a median load around 30,000 MW and IQR from 27,000 MW to 34,000 MW. Seasonal climate variation is also a key factor influencing electricity demand patterns, as evidenced by the highest median load in the season 3 (summer) box plot with the highest IQR of ~30,000–41,000 MW, while seasons 2 and 4 show comparatively lower and compact box plots with median loads of ~29,500 MW.



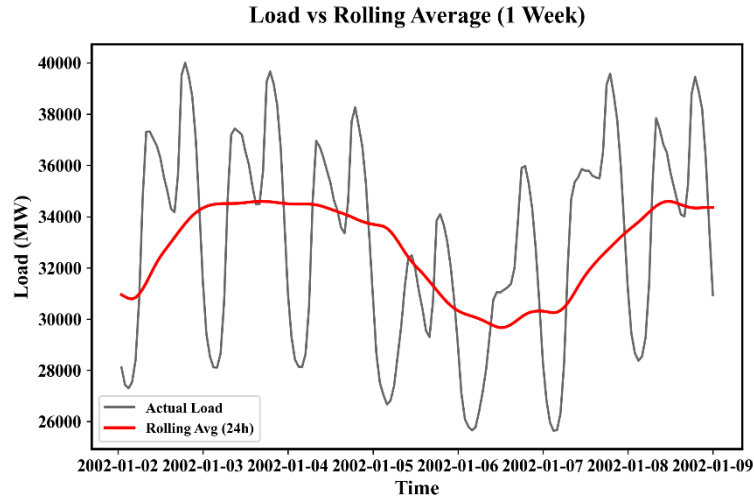
(a)



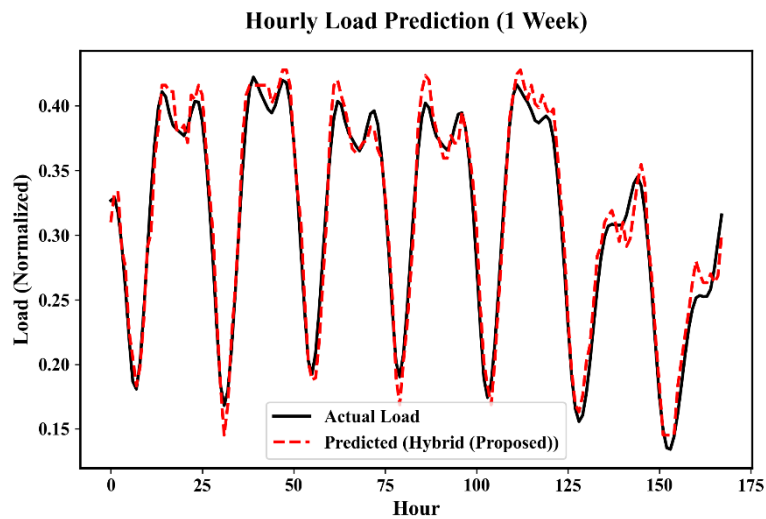
(b)

Figure 5: Load Demand Distribution (a) Hour of Day and (b) Day of Week

The electrical load demands have evident temporal patterns as shown in figure 5. The hourly plot indicates that load increased steadily from a low median of ~25,000 MW during early morning hours (1 AM - 5 AM) and peaked around 10 AM - 20:00 with ~33,000–35,000 MW due to the industrial and commercial activity during the day, before dropping at night after midnight. The day-of-week plot depicts consistently higher medians for the weekdays (days 0 to 4) of ~32,000 MW, which are followed by a significant decrease in electricity demand during the weekend (days 5 to 6) of ~29,500 MW, proving that not only intra-day but also weekly cycles play an important role in determining electricity demand.



(a)



(b)

Figure 6: (a) Load vs. Rolling Average and (b) Hourly Load Prediction Over One Week

The load demand trends and model prediction performance for 1-week period are shown in Figure 6. The rolling average plot is a 24-hour rolling average of the real load data (grey) showing the daily cycle with the average load fluctuating between ~25,000–40,000 MW, with the trend of the plot being a downward dip to ~29,500 MW mid-week (Jan 5–6) followed by a recovery to ~34,000 MW (red). The hourly prediction plot shows that the model's predictions (red dashed) closely match the actual normalized load (black) and follow the diurnal peaks and valleys in the load

between ~ 0.14 and ~ 0.42 with only minor differences at the sharp transitions, highlighting that the model has good temporal forecasting capabilities.

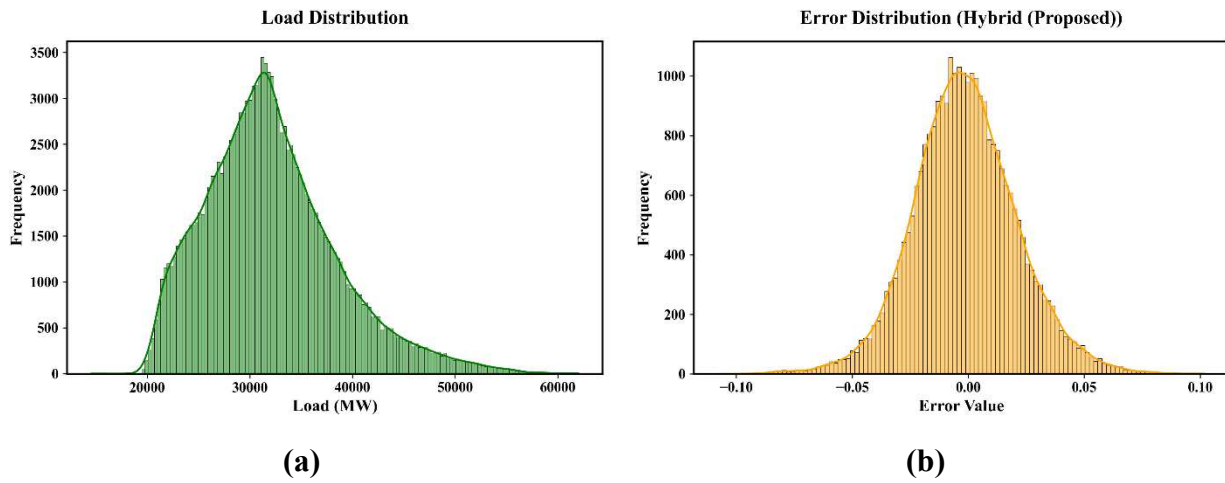


Figure 7: Error Distribution of (a) Load Distribution and (b) Hybrid proposed Model

Statistical features of the errors for the prediction of the Hybrid model and actual data are illustrated in Figure 7. As seen from the graph of load distribution, the distribution of hourly loads has strong skewness to the right with peak values in the interval between 25,000 MW and 35,000 MW with frequency of occurrence $> 3,400$ hourly loads, followed by a long tail towards 60,000 MW suggesting that there is sometimes high demand. The error distribution for the Hybrid model is bell-shaped and centered around zero, with most errors falling within the range of -0.05 to $+0.05$. This pattern is indicative of minimal systematic error and a well-balanced distribution of prediction residuals, confirming the model's reliability and predictive capability under different loading scenarios.

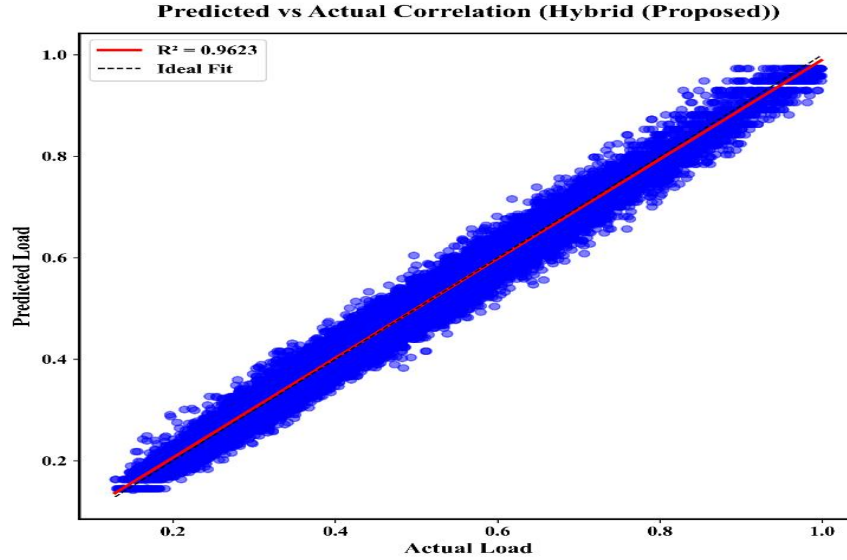


Figure 8: Predicted vs. Actual Correlation of the Proposed Model

From Figure 8, it is observed that the correlation between the actual and estimated normalized load values through the proposed Hybrid model is found to be $R^2 = 0.9623$. This data is clustered in a tight band around the regression line (red) that lies close to the ideal fit (black dashed diagonal) over the whole load range from 0.15 to 1.0. The high R^2 value validates the model by showing that about 96.23% of the actual load demand is accounted for, showing a strong predictive capability and excellent generalisation capabilities both at low and high demand.

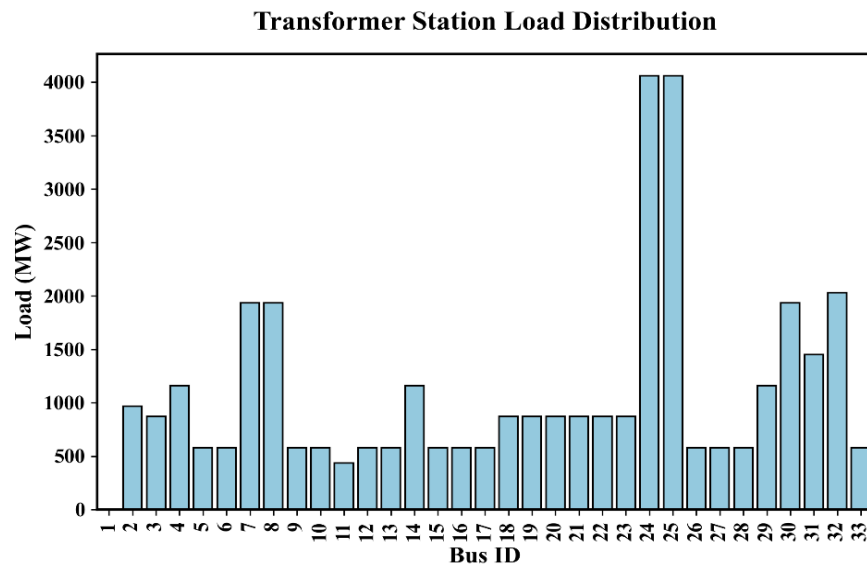


Figure 9: Transformer Station Load Distribution Across Bus IDs

The load distribution between 33 transformer station bus IDs is shown in figure 9, in which a large spread is seen in the load. Moderate loads are experienced by most of the buses ranging between 500 MW to 900 MW, but significant peaks are observed at Bus ID 7 and 8 (in the range of ~1950 MW); at Bus ID 14 (~1150 MW); at Bus ID 30 (~1950 MW) and at Bus ID 32 (~2050 MW); and most prominently at Bus ID 24 and 25 with loads of ~4050 MW each, indicating heavily loaded critical nodes. The figures for bus IDs 1, 27, and some others are very low or zero, indicating that these stations may have low or non-existent loads, underscoring the disparity in load distribution across the network.

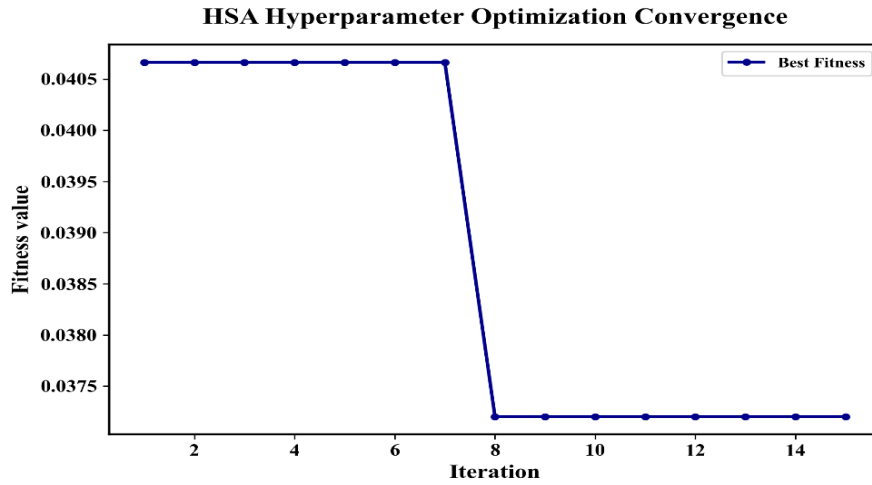
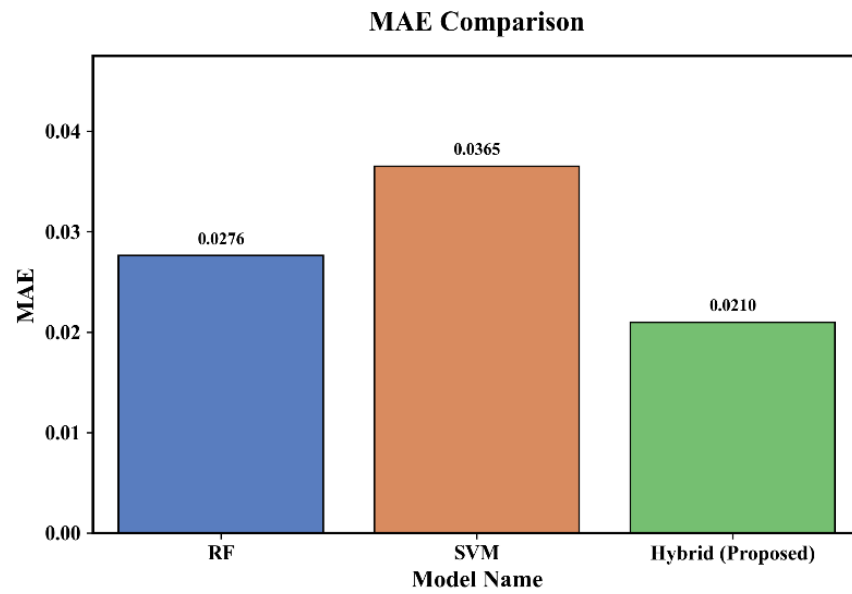


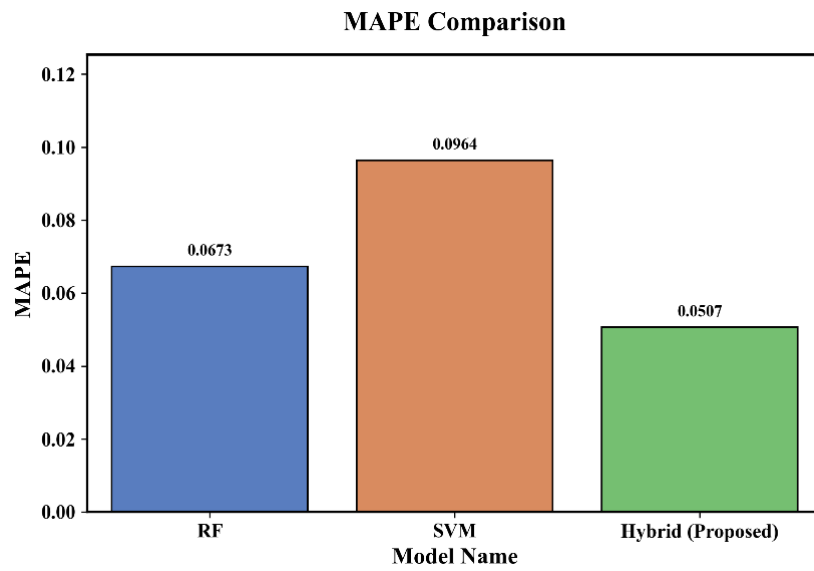
Figure 10: Convergence analysis of HSA-based hyperparameter optimization

Figure 10 represents the convergence of the HAS for hyperparameters tuning of the proposed Hybrid LSTM-XGB forecasting system. The fitness value is kept constant at around 0.0407 from iteration 1 to iteration 7, showing that the parameter is exploring the process of optimization. After iteration 8, there is a remarkable improvement in the fitness value, dropping to approximately 0.0372, indicating that an optimum combination of hyperparameters has been found. The fitness value does not change after iteration 8, which shows a convergence stability and optimization consistency. The fitness value has decreased from 0.0407 to 0.0372, which shows a better accuracy in forecasting and a less prediction error using HSA based parameter tuning.

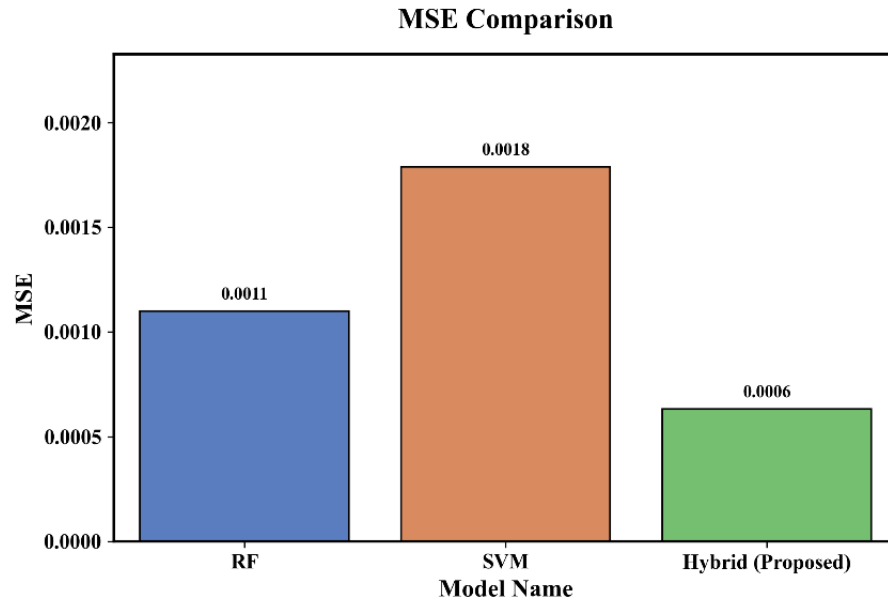
4.3 Analysis of comparative model performance metrics



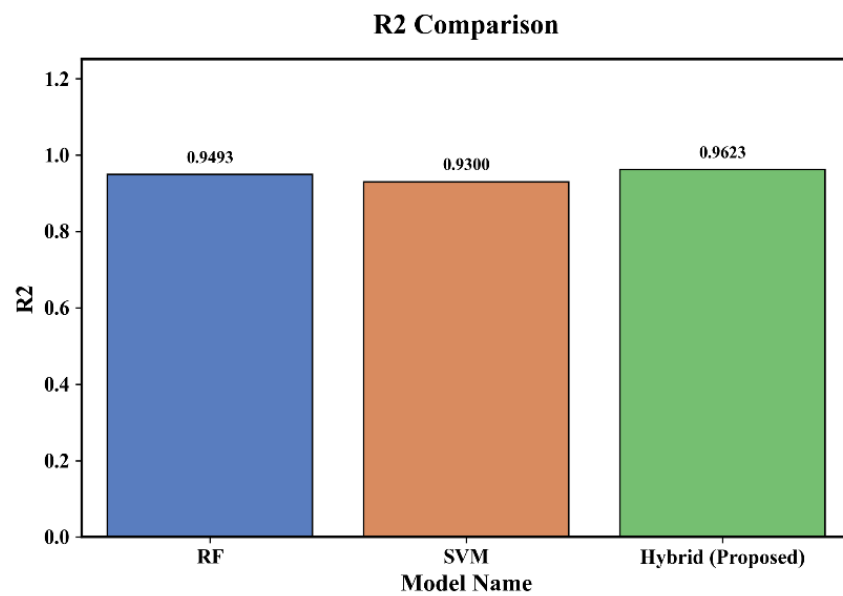
(a)



(b)



(c)



(d)

Figure 11: Performance comparison of Models Across (a) MAE, (b) MAPE, (c) MSE and (d) R^2 Metrics

The prediction error performance of the three models, namely RF, SVM, and the proposed Hybrid model, is illustrated with respect to MAE, MAPE, MSE, and R^2 as depicted in Figure 11. The error

values and R^2 of the Hybrid (Proposed) model are also found to be smallest among the other three models for all four metrics, implying it is the most accurate at predicting the values. The SVM model has the highest error rates among all the models on all metrics MAE (0.0365), MAPE (0.0964), MSE (0.0018), and R^2 (0.9300), showing relatively poor generalization performance. From the above results, the Hybrid model is seen to perform significantly better than the other two models in terms of reduction in MSE (approximately 45.5% and 66.7%), MAE (approximately 23.9% and 42.5%), MAPE (approximately 24.7% and 47.4%), and enhancement in R^2 (approximately 1.37% and 3.47%).

Table 3: Comparative performance analysis with existing works

Author	MAE	MAPE	MSE	R^2
(Ramesh et al., 2022)	0.0412	0.0985	0.0021	0.9124
(Chang et al., 2002)	0.0387	0.0914	0.0019	0.9187
(L’Heureux et al., 2022)	0.0315	0.0742	0.0013	0.9385
Proposed Hybrid LSTM–XGB + HSA	0.0199	0.048	0.0006	0.9638

Table 3 shows the comparison between the proposed Hybrid LSTM–XGB model and other forecasting models based on MAE, MSE, R^2 , and MAPE values. The proposed model achieved the best performance with MAE of 0.0199, MSE of 0.0006, and R^2 of 0.9638 and MAPE of 0.0480. In comparison, (Ramesh et al., 2022) obtained MAE of 0.0412 and R^2 of 0.9124, while Rung-Fang (Chang et al., 2002) achieved MAE of 0.0387 and R^2 of 0.9187. In a like manner (L’Heureux et al., 2022) reported better performance with MAE equal to 0.0315 and R^2 value equal to 0.9385, which is still less than the proposed framework.

4.4 Ablation study

Table 4: Ablation analysis of the proposed model

Configuration	MSE	MAE	MAPE	R^2
Proposed (Hybrid+HSA)	0.000578	0.019941	0.047968	0.963848
w/o XGBoost (LSTM)	0.00068	0.022084	0.051672	0.932939
w/o LSTM (XGB)	0.001564	0.034349	0.082082	0.903854
w/o HSA (Defaults)	0.000682	0.022334	0.052765	0.946498

w/o PCA	0.000791	0.024118	0.058904	0.941275
w/o CA	0.000845	0.025406	0.061237	0.938842
w/o Weather Features	0.000956	0.027881	0.067924	0.929615
w/o Topology Features	0.001021	0.029447	0.071302	0.924381

The results of the proposed forecasting framework have been tabulated in Table 4 .The complete Hybrid+HSA model had the highest R^2 value (0.963848) lowest MSE (0.000578), MAE (0.019941), MAPE (0.047968) and XGBoost had a minor impact on forecasting accuracy, while omission of LSTM had the greatest impact with the highest prediction error and lowest R^2 . In the same way, eliminating HSA optimization reduced the convergence quality and predictive stability.

4.5 Discussion

In order to improve the efficiency of the short-term load forecasting in distribution systems and transformer station areas, this paper presents an intelligent forecasting model based on the hybrid model, where the temporal behavior of load, weather, and other electrical properties are integrated into one forecasting framework with HSA-based hyper parameter tuning. The preprocessing stage ensured consistency of the data by interpolating, filtering, removing outliers using Z score, and normalizing it using Min–Max normalization, and CA and PCA helped eliminate redundant attributes and increase the efficiency of forecasting. The proposed model was proven to model the temporal variations in loads, the seasonal load patterns and the showing of the transformer operation patterns through the experimental analysis. The model achieved superior predictive performance with MAE of 0.0210, R^2 of 0.9623, MAPE of 0.0507, and MSE of 0.0006, outperforming RF and SVM models, where RF obtained MAE of 0.0276, MAPE of 0.0673, MSE of 0.0011, and R^2 of 0.9493, while SVM achieved MAE of 0.0365, MAPE of 0.0964, MSE of 0.0018, and R^2 of 0.9300. In addition, the ablation analysis verified the effectiveness of LSTM, XGBoost, and HSA; The overall Hybrid+HSA model achieved the best results with a mean squared error of 0.000578, mean absolute error of 0.019941, mean absolute percentage error of 0.047968, and coefficient of determination of 0.963848, which is enough to prove the robustness and forecasting capabilities of the proposed intelligent load forecasting framework.

5. Conclusion

This study proposed an intelligent hybrid forecasting model based on the HSA and hyperparameter optimization for the prediction of accurate load in the distribution systems and transformer station

area. The framework discussed the use of the advanced preprocessing, CA-based feature selection, PCA-based dimensionality reduction, and SW-based temporal sequence construction techniques to enhance the forecasting efficiency and stability of learning, which involved integrating historical load demand, weather information and power network operational information. The experimental results showed that the proposed model performed better than RF and SVM models in all the evaluation metrics with MAPE of 0.0507, MSE of 0.0006, R^2 of 0.9623 and MAE of 0.0210. Further, the load mapping of the transformer stations was effective in identifying the bus regions with heavy loads, which is useful for detecting and monitoring these overloaded bus regions in the operation of the transformer stations.

5.1 Limitation and Future work

The present work is primarily based on a short-term forecasting without real-time fault analysis and adaptive online learning. Moving forward, incorporating Graph Neural Networks, transformer-based models, handling uncertainty in renewable energy and incorporating information from IoT sensors and explainable AI would enhance the scalability, predictability of forecasts and real-world implementation of smart grids.

Dataset link:

1. PJM Hourly Energy Consumption Dataset :
<https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption/>
2. NOAA Climate Dataset : <https://www.kaggle.com/datasets/noaa/gsod/>
3. MATPOWER IEEE Bus Dataset : <https://matpower.org/download/>

Reference

- Bouhamed, O., Dissem, M., Amayri, M., & Bouguila, N. (2025). Transformer-based deep probabilistic network for load forecasting. *Engineering Applications of Artificial Intelligence*, 152, 110781. <https://doi.org/10.1016/j.engappai.2025.110781>
- Brahmachary, R., & Ahmed, I. (2025). Optimal distribution network expansion and EV charging station allocation based on load forecasted using machine learning. *IEEE Transactions on Industry Applications*. <https://doi.org/10.1109/TIA.2025.3546595>

- Chang, R.-F., Leou, R.-C., & Lu, C.-N. (2002). Distribution transformer load modeling using load research data. *IEEE Transactions on Power Delivery*, 17(2), 655–661. <https://doi.org/10.1109/61.997955>
- Fan, J., Zhuang, W., Xia, M., Fang, W., & Liu, J. (2024). Optimizing attention in a Transformer for multihorizon, multienergy load forecasting in integrated energy systems. *IEEE Transactions on Industrial Informatics*, 20(8), 10238–10248. <https://doi.org/10.1109/TII.2024.3362480>
- Fang, B., Xu, L., Luo, Y., Luo, Z., & Li, W. (2024). A method for short-term electric load forecasting based on the FMLP-iTransformer model. *Energy Reports*, 12, 3405–3411. <https://doi.org/10.1016/j.egyr.2024.02.145>
- Gao, S., et al. (2025). Short-term residential load forecasting via transfer learning and multi-attention fusion for EVs' coordinated charging. *International Journal of Electrical Power & Energy Systems*, 164, 110349. <https://doi.org/10.1016/j.ijepes.2025.110349>
- Gong, M., Zhao, Y., Sun, J., Han, C., Sun, G., & Yan, B. (2022). Load forecasting of district heating system based on Informer. *Energy*, 253, 124179. <https://doi.org/10.1016/j.energy.2022.124179>
- Gorginpour, H., Ghimatgar, H., & Toulabi, M. S. (2022). Lifetime estimation and optimal maintenance scheduling of urban oil-immersed distribution-transformers considering weather-dependent intelligent load model and unbalanced loading. *IEEE Transactions on Power Delivery*, 37(5), 4154–4165. <https://doi.org/10.1109/TPWRD.2021.3139854>
- Grabner, M., Wang, Y., Wen, Q., Blažič, B., & Štruc, V. (2023). A global modeling framework for load forecasting in distribution networks. *IEEE Transactions on Smart Grid*, 14(6), 4927–4941. <https://doi.org/10.1109/TSG.2023.3266655>
- He, C., Zhu, J., Lan, J., Li, S., Wu, W., & Zhu, H. (2022). Optimal planning of electric vehicle battery centralized charging station based on EV load forecasting. *IEEE Transactions on Industry Applications*, 58(5), 6557–6575. <https://doi.org/10.1109/TIA.2022.3185277>
- He, Z., et al. (2023). Short-term power load forecasting of multi-charging piles based on improved gate recurrent unit. *IEEE Access*, 12, 2490–2499. <https://doi.org/10.1109/ACCESS.2023.3342786>

- Huang, Y., Wu, S., Wang, Z., Liu, X., Li, C., & Hu, Y. (2025). Causality-aware multi-graph convolutional networks with critical node dynamics for electric vehicle charging station load forecasting. *IEEE Transactions on Smart Grid*. <https://doi.org/10.1109/TSG.2025.3570955>
- Hussain, A., Eswarakrishnan, V., Aslam, A., & Tripura, S. (2025). Charging stations demand forecasting using LSTM based hybrid Transformer model. *Scientific Reports*, *15*(1), 36639. <https://doi.org/10.1038/s41598-025-90657-9>
- L'Heureux, A., Grolinger, K., & Capretz, M. A. M. (2022). Transformer-based model for electrical load forecasting. *Energies*, *15*(14), 4993. <https://doi.org/10.3390/en15144993>
- Liu, G., et al. (2026). An integrated method for electric load forecasting in large public buildings based on load patterns clustering and multi-scale temporal feature selection. *Energy*. <https://doi.org/10.1016/j.energy.2026.140438>
- Mogos, A. S., Liang, X., & Chung, C. Y. (2023). Distribution transformer failure prediction for predictive maintenance using hybrid one-class deep SVDD classification and lightning strike failures data. *IEEE Transactions on Power Delivery*, *38*(5), 3250–3261. <https://doi.org/10.1109/TPWRD.2022.3224320>
- Mulla, R. (2018). *PJM hourly energy consumption dataset* [Data set]. Kaggle. Kaggle dataset
- National Oceanic and Atmospheric Administration. (2024). *Global surface summary of the day (GSOD) climate dataset* [Data set]. Kaggle. Kaggle GSOD dataset
- Özen, S., Yazıcı, A., & Atalay, V. (2025). Hybrid deep learning models with data fusion approach for electricity load forecasting. *Expert Systems*, *42*(2), e13741. <https://doi.org/10.1111/exsy.13741>
- Pentsos, V., Tragoudas, S., Wibbenmeyer, J., & Khdeer, N. (2025). A hybrid LSTM-Transformer model for power load forecasting. *IEEE Transactions on Smart Grid*. <https://doi.org/10.1109/TSG.2025.3535407>
- Power Systems Engineering Research Center. (2025). *MATPOWER IEEE bus test system dataset* [Data set]. MATPOWER. MATPOWER download page

- Ramesh, J., Shahriar, S., Al-Ali, A. R., Osman, A., & Shaaban, M. F. (2022). Machine learning approach for smart distribution transformers load monitoring and management system. *Energies*, 15(21), 7981. <https://doi.org/10.3390/en15217981>
- Syed, D., et al. (2021). Deep learning-based short-term load forecasting approach in smart grid with clustering and consumption pattern recognition. *IEEE Access*, 9, 54992–55008. <https://doi.org/10.1109/ACCESS.2021.3071209>
- Tian, R., Wang, J., Sun, Z., Wu, J., Lu, X., & Chang, L. (2025). Multi-scale spatial-temporal graph attention network for charging station load prediction. *IEEE Access*, 13, 29000–29017. <https://doi.org/10.1109/ACCESS.2025.3541118>
- Torres-Bermeo, P., López-Eugenio, K., Del-Valle-Soto, C., Palacios-Navarro, G., & Varela-Aldás, J. (2025). Sizing and characterization of load curves of distribution transformers using clustering and predictive machine learning models. *Energies*, 18(7), 1832. <https://doi.org/10.3390/en18071832>
- Wang, K., Zhang, J., Li, X., & Zhang, Y. (2023). Long-term power load forecasting using LSTM-Informer with ensemble learning. *Electronics*, 12(10), 2175. <https://doi.org/10.3390/electronics12102175>
- Zabin, R., Haque, K. F., & Abdelgawad, A. (2024). PredXGBR: A machine learning framework for short-term electrical load prediction. *Electronics*, 13(22), 4521. <https://doi.org/10.3390/electronics13224521>
- Zhan, X., Kou, L., Xue, M., Zhang, J., & Zhou, L. (2022). Reliable long-term energy load trend prediction model for smart grid using hierarchical decomposition self-attention network. *IEEE Transactions on Reliability*, 72(2), 609–621. <https://doi.org/10.1109/TR.2022.3149571>
- Zhang, Y., et al. (2021). A prediction model of hot spot temperature for split-windings traction transformer considering the load characteristics. *IEEE Access*, 9, 22605–22615. <https://doi.org/10.1109/ACCESS.2021.3056975>
- Zhang, Y., Liu, C., Rao, X., Zhang, X., & Zhou, Y. (2024). Spatial-temporal load forecasting of electric vehicle charging stations based on graph neural network. *Journal of Intelligent & Fuzzy Systems*, 46(1), 821–836. <https://doi.org/10.3233/JIFS-231775>

- Zhao, H., Wu, Y., Ma, L., & Pan, S. (2023). Spatial and temporal attention-enabled Transformer network for multivariate short-term residential load forecasting. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–11. <https://doi.org/10.1109/TIM.2023.3305655>
- Zhao, Z., et al. (2021). Short-term load forecasting based on the Transformer model. *Information*, 12(12), 516. <https://doi.org/10.3390/info12120516>
- Zhuang, W., Fan, J., Xia, M., & Zhu, K. (2023). A multi-scale spatial–temporal graph neural network-based method of multienergy load forecasting in integrated energy system. *IEEE Transactions on Smart Grid*, 15(3), 2652–2666. <https://doi.org/10.1109/TSG.2023.3332178>