Zaria Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria – Nigeria. Vol. 14 No. 1 & 2, September 2025. ISSN: 0261 – 1570.

Electric Load Forecasting using Machine Learning and Traditional Models: A Review

A. U,Lawan[1]
[1] Department of Electrical Engineering, Bayero university Kano-Nigeria.
[1]*e-mail:aulawan.ele@buk.edu.ng

Keywords: –
Machine learning;
Electric load forecasting;
Modeling electricity loads;
Methods and models of forecasting;
Short term Load Forecasting

Article History: –
Received: 07 May, 2025.
Reviewed: 12 May, 2025
Accepted: 30 May, 2025
Published: 02 June, 2025

ABSTRACT

Electric load forecasting (ELF) is an essential procedure in the electricity industry's planning, with a significant impact on electric capacity scheduling and power systems management. As a result, it has garnered growing attention from the academic community. Load forecasting has become an essential component of electricity utility firms. In order to ensure uninterrupted and reliable power supply to consumers, dlecision-makers in the utility sector need to accurately predict the future electricity demand, minimizing any margin of error. Therefore, the precision of electric load forecasting is crucial for scheduling energy generation capacity and managing power systems. This paper examines the techniques and a model used for predicting electricity load, and provides an overview of the latest advancements in electric load forecasting technologies. It focuses on recent studies that have explored the combination of multiple machine learning algorithms to create hybrid models. A total of 44 academic articles were utilized to compare various projects based on specific criteria, including the time frame, project size, and the methods and models implemented.

1. INTRODUCTION

The contemporary power system necessitates a continuous provision of electricity to the load side. It is necessary to have a precise understanding of forecasting current and future load demand with little margin of error. Electric load/demand forecasting is an essential procedure in the strategic planning of the electricity sector and has a pivotal function in the operation of electric power networks. (Bian et al., 2022). The electric power demand projection is heavily tied to the economy's development, and it is also related to national security and the daily operation of society. Therefore, the accuracy of electric load forecasting has tremendous importance for energy generating capacity scheduling and power system management, since these accurate forecasts lead to large savings in operating and maintenance costs, and right decisions for future development. Furthermore, electric power load forecasting represents the starting step in building future production, transmission, and distribution infrastructure.

However, the accuracy of electric load forecasting cannot often satisfy our intended result since it is influenced by several unknown and uncontrollable aspects such as economic development, human social activities, country regulations, and climatic change. So far, there is no clear criterion for identifying the range of load projections. However, load forecasting has been separated in terms of the prediction period into three groups; short-term forecasts, medium-term forecasts, and long-term predictions.

It is of tremendous practical value to examine the load forecasting under the coupling conditions of numerous loads of integrated energy systems. (Hammad et al., 2020). In this context, it is vital to maintain track of the newest research progress of load forecasting methods and grasp the current research hotspots and directions of load forecasting for the development and construction of integrated energy systems.

2. METHODOLOGY

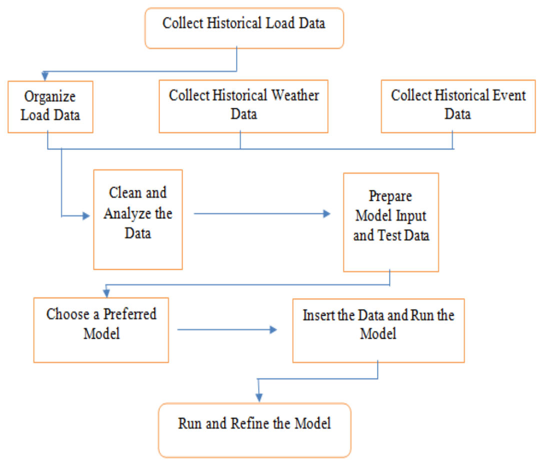

Figure 1: Flow chart of building a load forecasting model

2.1 Review of Some Methods and Models Employed in Electric Load Forecasting

This subsection presents some of the major techniques (methods and models) that have been considered for electric load forecasting that are currently available in literature. The techniques are:

2.1.1. Gradient Boosting Regression (GBR)

Gradient Boosting is a machine learning model used for both regression and classification tasks, creating a prediction model by combining multiple weak models, often decision trees. Similar to other boosting methods, it builds the model progressively, but it extends this concept by allowing the optimization of any differentiable loss function. The core idea behind Gradient Boosting Regression (GBR) is that boosting can be seen as an optimization process applied to a specific cost function. The algorithm works by iteratively selecting a weak model (a hypothesis) that moves in the direction of the negative gradient, optimizing the cost function over a function space.

This interpretation of boosting as a functional gradient method has led to the development of boosting algorithms not only in regression and classification but also in other areas of machine learning and statistics (Yamasaki et al., 2024).

2.2.2 k-Nearest Neighbors (kNN)

The k-Nearest Neighbors (kNN) algorithm predicts a target value by performing local interpolation using the target values of the closest neighbors from the training dataset. These neighbors' contributions are influenced by their Euclidean distance from the query sample—the closer the neighbor, the greater its impact. kNN is widely applied in various domains, including recommendation systems, financial market forecasting, and text classification, among others (Yamasaki et al., 2024).

2.2.3 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a supervised learning model that can address both regression and classification tasks. The Support Vector Machine (SVM) training algorithm constructs a model to classify new examples into one of two categories, functioning as a binary linear classifier without relying on probabilistic methods. SVM represents training examples as points in space, aiming to maximize the margin between the two categories. When new examples are introduced, they are placed within this space and classified based on their position relative to the margin. Beyond linear classification, SVM can also perform non-linear classification effectively by utilizing the kernel trick, which implicitly maps input data into high-dimensional feature spaces (Yamasaki et al., 2024).

2.2.4 Convolutional Neural Network

Convolutional Neural Network (CNN) is a specialized deep learning model commonly used for tasks such as classification, pattern recognition, forecasting, and regression. Its architecture comprises convolutional layers, activation functions, pooling layers, fully connected layers, and a classification layer. This structured, layer-by-layer design enhances its ability to extract features from input data, making it more effective than other deep learning models. Each layer in a CNN has a distinct role. Convolutional layers use filters to extract feature maps that identify and describe input data features. Different filters are applied at each layer to generate unique feature maps. The Rectified Linear Unit (ReLU) activation function is applied to these outputs to perform the activation process. Pooling layers in the CNN aggregate features extracted from the convolutional layers. Two common pooling modes are average pooling, which computes the average of the features, and max-pooling, which selects the maximum feature value. The output of each pooling layer serves as the input to the next convolutional layer. Fully connected layers, acting as feed-forward neural networks, determine the weights and biases of the data. The feature map from the last pooling layer becomes the input to these fully connected layers. Finally, the SoftMax layer produces the network's output after training is completed (Moradzadeh et al., 2022).

2.2.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks were developed to enhance the performance of Recurrent Neural Networks (RNNs), addressing issues such as vanishing and exploding gradients. LSTMs are widely used for tasks such as estimation, regression, nonlinear data modeling, and categorization. The LSTM architecture consists of three gates: the input gate (i_t), forget gate (f_t), and output gate (o_t), along with a memory cell. The memory cell manages information updates, with the input gate determining what to add and the forget gate deciding what to remove. The output gate regulates the flow of information from the input, memory cell, and the hidden state. The activation process in LSTM is carried out using the sigmoid activation function (sigma). Each gate has its own weight matrix, represented as (W_i), (W_f), (W_o), and (W_c), and corresponding biases (b_i), (b_f), (b_o), and (b_c). In this architecture, ($h_{\{t-1\}}$) denotes the hidden state from the previous time step, (x_t) represents the input at the current time step, (c_t) is the current memory cell state, and (h_t) is the current hidden state. Long Short-Term Memory (LSTM) networks were developed to enhance the performance of Recurrent Neural Networks (RNNs), addressing issues such as vanishing and exploding gradients. LSTMs are widely used for tasks such as estimation, regression, nonlinear data modeling, and categorization (Moradzadeh et al., 2022).

2.2.6 Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a newer type of neural network model based on LSTM. Unlike LSTM, GRU features a simpler structure, fewer parameters, and shorter training times, often achieving better performance on certain tasks. The internal operations of the GRU are described by the following equation:

$$z_t = \sigma(W_z.[h_{t-1}, x_t]) \quad (1)$$

$$r_t = \sigma(W_r.[h_{t-1}, x_t]) \quad (2)$$

$$h'_t = \tanh(W.[r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t \quad (4)$$

where: z_t represents the update gate, r_t is the reset gate, σ sigma is the sigmoid function, and tanh is the hyperbolic tangent function. W_r, W_z, and W are the parameter training matrices. The candidate hidden state is denoted as h'_t, while h_t represents the hidden state that is passed to the next time step, and h_{t-1} is the hidden state from the previous time step. Finally, x_t refers to the current input data (Fan et al., 2024).

2.2.7 ForecastNet

ForecastNet is a deep learning framework introduced in 2020 by the Joel Janek Dabrowski team. It utilizes a deep feedforward architecture to create time-varying models. As a type of deep learning model, a deep feedforward network, or multilayer perceptron (MLP), consists of n_I input nodes, n_O output nodes, and a sequence of hidden layers. When the input is x = $x_{t-nI+1:t}$, ForecastNet predicts the output $x'_{t+1:t+4}$. The hidden layers in ForecastNet employ a feedforward neural network structure, with interleaved outputs to address the vanishing gradient problem. In deep learning, the repeated application of the chain rule during gradient calculation can result in gradient vanishing. By using interleaved outputs, ForecastNet decomposes the gradient chain into the sum of multiple terms, making it more stable than a product of factors and effectively reducing the network depth (Li et al., 2022).

2.2.8 Extreme gradient Boost (XGBoost)

XGBoost works by combining multiple weak classifiers (Tree 1-t) into a strong classifier to enhance prediction accuracy. In regression tasks, for instance, each tree (except the first) is trained to minimize the loss (negative gradient) of the cumulative results from all preceding trees. By iteratively adding the loss to the prior predictions, the model progressively approaches the true values.

XGBoost offers several advantages:

1. It improves prediction accuracy through the sequential integration of decision trees, demonstrating strong learning capabilities.
2. It uses regularization to manage model complexity, preventing overfitting and enhancing generalization.
3. It speeds up optimization by applying second-order Taylor expansion to the loss function.

Overall, XGBoost is a highly accurate algorithm with robust learning capabilities, strong generalization, and resistance to overfitting. However, its accuracy improvements rely on parameter constraints and iterative processes (Deng et al., 2022).

3 COMPREHENSIVE DESCRIPTION OF ELECTRIC LOAD FORECASTING MODELS/METHODS

This subsection provides a detailed technical overview of the models and methods reviewed in this paper, categorized by their computational approach (machine learning vs. classical methods) and forecasting objectives (short-term vs. long-term).

3.1 Machine Learning (ML) Models

3.1.1 Gradient Boosting Machines (GBM, XGBoost, LightGBM)

Algorithm: Ensemble of weak learners (decision trees) trained sequentially to correct residuals. Its key features are:

I. XGBoost: Uses regularization (L1/L2) and parallel tree construction.
II. LightGBM: Optimized for speed via histogram-based splitting.

Strengths: Handles non-linearity, missing data, and feature importance.

Weaknesses: Struggles with long-term dependencies; hyperparameter-sensitive.

Performance: MAE: 1.5–4.0%, RMSE:2.0–5.0% (Deng et al., 2022).

3.1.2 k-Nearest Neighbors (kNN)

Algorithm: Predicts based on average of k most similar historical instances (Euclidean distance). Key Feature: Lazy learner (no training phase).

Strengths: Simple, no assumptions about data distribution.

Weaknesses: Computationally heavy for large datasets; sensitive to noise.

Performance: MAE: 3.0–7.0% (Yamasaki et al., 2024).

3.1.3 Support Vector Regression (SVR)

Algorithm: Maps data to high-dimensional space using kernels (RBF, polynomial) and fits a hyperplane.

Key Features: Maximizes margin while tolerating errors (ε-insensitive loss).

Strengths: Robust to outliers; effective in high dimensions.

Weaknesses: Poor scalability; kernel selection is critical.

Performance: MAE: 2.0–5.0% (Yamasaki et al., 2024).

3.3.4 Deep Learning Models

3.3.4.1 Long Short-Term Memory (LSTM)

Architecture: Recurrent network with input, forget, and output gates to manage memory.

Key Features: Addresses vanishing gradients; ideal for sequential data.

Strengths: Captures long-term dependencies (e.g., daily/weekly cycles).

Weaknesses: Computationally expensive; requires large data.
Performance: MAE: 0.8–2.5% (Moradzadeh et al., 2022).
3.3.4.2 Gated Recurrent Unit (GRU)
Architecture: Simplified LSTM with reset and update gates (fewer parameters).
Key Features: Faster training than LSTM but slightly less accurate.
Performance: MAE: 1.0–3.0% (Fan et al., 2024).
3.3.4.3 Convolutional Neural Networks (CNN)
Architecture: Uses 1D convolution to extract local patterns (e.g., daily load shapes).
Key Features: Pooling layers reduce dimensionality.
Strengths: Effective for spatial-temporal data.
Weaknesses: Lacks native sequential modeling (often paired with LSTM).
Performance: MAE: 1.0–3.0% (Chen et al., 2023).
3.3.4.4 Hybrid Models (TCN-LSTM, CNN-GRU)
Architecture: Combines CNN (spatial features) with LSTM/GRU (temporal features).
Key Features: TCN (Temporal Convolutional Network) uses dilated convolutions for long sequences.
Strengths: State-of-the-art accuracy for short-term forecasts.
Weaknesses: Complex deployment.
Performance: MAE: 0.7–2.0% (Zhang et al., 2023).
3.3.4.5 Transformers Architecture: Self-attention mechanisms to weigh input importance.
Key Features: Positional encoding for time-series; parallelizable.
Strengths: Handles long-range dependencies better than RNNs.
Weaknesses: Data-hungry; lacks interpretability.
Performance: Emerging use (limited metrics in reviewed papers).
4.3.4.6 Federated Learning (FedForecast)
Algorithm: Decentralized training across edge devices (e.g., smart meters).
Key Features: Preserves privacy; aggregates model updates (not raw data).
Strengths: Scalable for distributed grids.
Weaknesses: Communication overhead; potential bias.
Performance: MAE: 1.2–3.5% (Liu et al., 2023).

3.2. Classical (Non-ML) Methods**

1. Autoregressive Models (ARIMA, SARIMA)
Algorithm: ARIMA: AutoRegressive (AR) + Integrated (I) + Moving Average (MA).
2. SARIMA: Adds seasonal differencing.
Key Features: Linear assumptions; requires stationarity.
Strengths: Interpretable; works with small data.
Weaknesses: Fails with non-linear trends (e.g., renewable integration).
Performance: MAE: 3.0–8.0% (Hammad et al., 2020).
3 Decomposition Techniques
a) STL (Seasonal-Trend Decomposition via Loess)
Algorithm: Separates time-series into trend, seasonal, and residual components.
Use Case: Preprocessing step for ML models.
Performance: Reduces LSTM MAE by 10–20% when combined (Yamasaki et al., 2024).
b) Empirical Mode Decomposition (EMD)
Algorithm: Data-driven decomposition into Intrinsic Mode Functions (IMFs).
Variants: CEEMDAN (noise-assisted) improves stability.
Weaknesses: Mode mixing issues.
4 Physics-Based Models
Approach: Uses differential equations for grid dynamics (e.g., thermal inertia).
Strengths: No training data needed.
Weaknesses: Inflexible to external factors (e.g., weather).
Performance:MAE:4.0–10.0%(Bianet al., 2022).
5 Fuzzy Logic Systems
Algorithm: Rule-based inference (e.g., "IF temperature is high, THEN load increases").
Strengths: Handles imprecise inputs.
Weaknesses: Rules require expert knowledge.
Performance: MAE: 3.5–8.5% (Zulfiqar et al., 2023).

4. LITERATURE REVIEW

Forecasting energy can be grouped into three main categories: short, medium, and long-term. There are two basic ways for carrying out energy load forecasting: physics-based models and statistical/machine learning based models. Physics based models incorporate engineering principles which relay on a complex combination of material, structural, and geometric aspects. The statistical based models examine previous energy consumption data, by implementing ML algorithms to quantitatively represent the link between the historical data and variables affecting energy consumption.

(Zhang et al., 2023) Proposed model was a novel collaborative TCN-LSTM-MTL model for short-term load forecasting. It utilizes mobility data, temporal convolutional networks, and multi-task learning to enhance its performance.

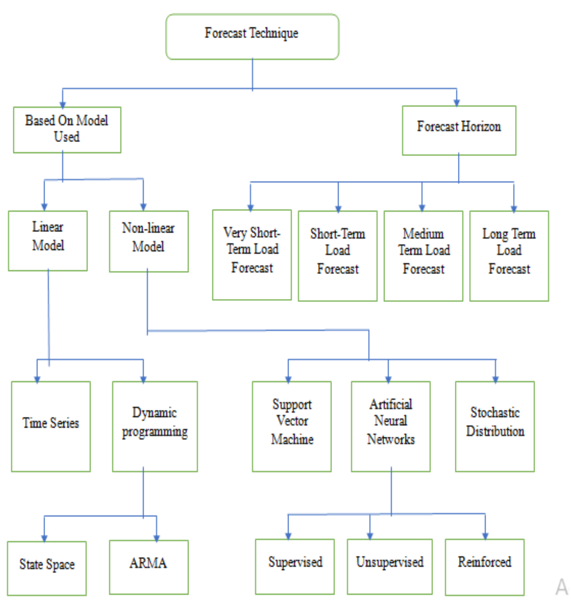

Figure 2: Classification of load forecasting techniques

The incorporation of parameter sharing layers and the implementation of a structure with residual convolution enhances the data input diversity of the forecasting model which allows the model to have a broader time series receptive field. (Liu et al., 2023) Presented a distributed short-term individual load forecasting method called FedForecast, which was built on the federated learning framework. This method ensures the privacy of users and maximizes the utilization of edge computing resources. Within this paradigm, the transmission of forecasting models takes precedence over the transmission of load data during the training of the models. Also, a novel short term residential load forecasting method was proposed in (Su et al., 2023) which considers the presence of uncertainties from several sources, a divide-and-conquer technique was suggested to address these uncertainties. This mechanism involves categorizing the residential load based on the device level and forecasting them separately for each source of uncertainty. Also, a multi-attribute adversarial learning technique, known as conditional Wasserstein generative adversarial network with gradient penalty (cWGAN-GP), was developed to account for the various contributing variables. (Yamasaki et al., 2024) Introduced a new hybrid machine learning model that integrated Gradient Boosting Regressor (GBR), Extreme Gradient Boosting (XGBoost), k-Nearest Neighbors (kNN), and Support Vector Regression (SVR). The model was evaluated both alone and in combination with signal decomposition techniques such as Seasonal and Trend decomposition using Loess (STL) and Empirical Mode Decomposition (EMD). EEMD, CEEMDAN, and EWT. Through Automated Machine Learning (AutoML), these models were integrated and their hyperparameters adjusted, predicting each load signal component using data from two sources: The National Operator of Electric System (ONS) and the Independent System Operators New England (ISO-NE), thereby boosting predictive capacity. A multi-stage integrated model based on decomposition, error factors, and a multi-objective evolutionary algorithm based on decomposition (MOEA/D) was proposed in (Fan et al., 2024). The proposed model consists of three stages: in the first stage, the gated recurrent unit (GRU) was used to predict the components of complete ensemble empirical modal decomposition with adaptive noise, and new data sets were obtained by combining them with the original data sets to fully mine the data characteristics. In the second stage, the MOEA/D based on angle and distance selection strategy and adaptive population generation strategy was used to optimize GRU network parameters with accuracy and diversity as the objective functions, obtaining several load forecasting models and error forecasting models that consider accuracy and diversity. In the third stage, a new nonlinear integration method based on GRU optimized by MOEA/D was utilized to integrate load forecasting values and error forecasting values, incorporating error factors to further increase forecasting accuracy. A day-ahead load forecasting approach that employs the smart meter data aggregated by residential customer's power consumption characteristics was introduced in (Han et al., 2023). First, the long-term trend information and daily variation information were derived from the residential load time series. According to the load characteristics shown by the daily load fluctuation statistics, the residential consumers were clustered into multiple groups using the K-means algorithm. The non-linear autoregressive neural network was used to forecast each cluster of users to capture their individual load patterns. Finally, the aggregated load at the system level was obtained by merging each cluster's predicting findings. (Pinheiro et al., 2023) Suggested a systematic approach from system level to low voltage considering not only the performance of the models but also the application, interpretability, and reproducibility of the method/model. An initial benchmark model was tested to improved regression models upgraded by integrating new synthetic explanatory variables and a different regression approach, generalized additive models. The error was reduced by 42%–47% relative to the benchmark model, keeping the interpretability. Additionally, a basic ensemble method was assessed to determine how it increases accuracy for specific periods, such as weekends, summer vacation, or public holidays, in which modeling was particularly difficult using solo generalized additive models. (Zulfiqar et al., 2023) developed a fast and accurate hybrid load forecasting model the developed model includes a locally weighted support vector regression (LWSVR) based forecaster with two modules. These modules were feature engineering (FE) and adaptive grasshopper optimization (AGO) based optimizers. A new methodology was proposed in (Tziolis et al., 2023) for direct short-term NLF at the distribution level, employing a Bayesian neural network model. The suggested model was optimized with decision heuristics based on a statistical post-processing step (i.e., clustering of daily irradiance patterns) for enhanced performance. To forecast electricity power net-load in renewable energy systems even though the presence of huge nonlinear variations makes its estimation tough and complicated, a straightforward strategy was provided in (Mokarram et al., 2023) to forecast signals with volatile characteristics to suit this demand. This framework blends deep learning of the multi-input LSTM network type with fuzzy system and discrete wavelet transforms. Wavelet-based transforms provide insight into hidden details and aid in forecasting places with significant chaos. Furthermore, technical indicators provide a technique to determine the trend and momentum of data and to select the ideal time frame for estimation. (Jalali et al., 2022) Presented a novel neuroevolutionary algorithm for handling the uncertainty involved with load predictions. In that paper, a novel modified evolutionary technique was proposed which was used to identify the best hyperparameters of 1D-Convolutional neural network (CNN). The probabilistic forecasts were formed by minimizing the mean scaled interval score loss function at 50%, 90% and 95% prediction intervals. The proposed neuroevolutionary method was tested on a global energy forecasting competition (GEFCom-2014) load dataset, and two separate experiments were conducted considering load only and one with load and temperature. (Smyl et al., 2024) Proposed a new short-term load forecasting (STLF) model based on contextually enhanced hybrid and hierarchical architecture integrating exponential smoothing (ES) and a recurrent neural network (RNN). The model was made of two simultaneously trained tracks, the context track and the main track. The background track offers additional information to the main track. It was taken from representative series and dynamically adjusted to adjust to the individual series projected by the main track. Shifting pattern of consumption by clients, called as concept drift, while the drift magnitude in load forecasting problems might vary dramatically over time, previous literature frequently assumes a fixed drift magnitude threshold, which should be dynamically modified rather than fixed during system evolution. (Bayram et al., 2023) offered a dynamic drift-adaptive Long Short-Term Memory (DA-LSTM) framework that can improve the performance of load forecasting models without requiring a drift threshold setting. (Bayram et al., 2023) Integrated many solutions into the framework based on active and passive adaptation approaches. (Aseeri et al., 2023) Provided a carefully-engineered forecasting approach for day-ahead electric power load forecasts assessed utilizing the European Network of Transmission System Operators for Electricity (ENTSO-E). Two steps were taken to configure the appropriate forecasting methodology: Firstly, a straightforward processing workflow was proposed to enable systematic preprocessing of raw multivariate time-discrete power data extracted from the ENTSO-E repository, including a stride-based sliding window approach to build time series-based batches appropriate for the supervised learning operation. Secondly, the lightweight form of recurrent neural network approach, namely gated recurrent units (GRU), was selected and properly calibrated to generate accurate multi-step forecasts, which was trained using the preprocessed multivariate time series data to render day ahead power load forecasts. Many unpredictable aspects to the planning and distribution of the power grid have been brought by connecting to the distributed power grid and rising active loads. To acquire more accurate and complete information about power load forecasting value, a short-term power load-interval multi-step forecasting approach based on ForecastNet was proposed in (Li et al., 2022). Firstly, single variable historical load data was utilized as input. Secondly, ForecastNet's deep feedforward architecture was presented to exactly capture the time-varying properties of load. Long-Short-Term-Memory (LSTM) model with simplex optimizer was developed in (Li et al., 2022) to forecast the electric load for a company during the COVID-19 epidemic. The forecasting method comprises of data processing, LSTM network creation and optimization. Firstly, some data processing procedures comprises information quantification, electric load data cleansing, correlation-coefficient-based medical data filtering, clustering-based medical data and electric load data filling. Then LSTM-based electric load forecasting model of enterprise was built during the COVID-19 epidemic. The LSTM network was trained and parameters were improved via simplex optimizer. In (Xiao et al., 2022) Probability forecast models were proposed to gather the uncertain information in the load power. Firstly, some error information was derived from the deterministic forecasting outcomes of point forecasting; secondly, the interval of time series data was divided according to the deterministic error information; finally, the Bootstrap method was used to estimate the confidence interval of the deterministic error information to acquire the uncertainty information in the power load data. The instability and randomness of the deterministic error were obtained by combining the interval forecasting approach so as to increase the accuracy of power load forecasting. Therefore, probability forecasting was paired with point forecasting to produce more accurate results. In another study (Xiang et al., 2022) a medium- and long-term power load prediction approach was proposed based on the two-layer categorical boosting (CatBoost) algorithm with multi-dimensional feature considerations. Simultaneously, the effects of economic fluctuation, power generation interruption, and meteorological data on power load were examined, wherein the dimension of power-load forecasting data features was broadened. (Deng et al., 2022) Proposed the Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model, which can warn the time period and detailed value of peak demand in advance. Firstly, based on Extreme Gradient Boosting (XGBoost) algorithm, the idea of Bagging was proposed to lower the output variance and boost the generalization capabilities of the method. Then, the mutual information (MI) between weather affecting elements and load was assessed to alter the input weight of the model and improve its ability to track weather changes. Next, considering the load, weather and time parameters, the extreme weather detection model was built to determine the occurrence range of peak load. The accuracy of the existing electric load forecasting systems relies on data quality due to noisy real-world situations, and data integrity due to hostile cyber-attacks. (Moradzadeh et al., 2022) presented a cyber-secure deep learning methodology that accurately anticipates electric load in power systems for a time horizon stretching from an hour to a week. The proposed deep learning framework systematically incorporated Autoencoder (AE), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) models (AE-CLSTM). In the operation of industrial energy systems, precise forecast of electric loads is a prerequisite to allow industrial users adjust their electric load dispatch and enhance energy efficiency. Therefore, (Zhu et al., 2023) offered a day ahead industrial load forecasting model applying load change rate features and merging the firefly algorithm to optimize the extreme learning machine and adaptive boosting algorithm (LCR-AdaBoost FA-ELM). The purpose of (Falces et al., 2023) was to analyze the difference in the net load forecasting error achieved by models using or not using behind-the-meter PV generation data. The PV plant was connected to the lower voltage side of the electricity substation, representing a penetration level of over 35% of the total load. (Falces et al., 2023) revealed that the best forecasting results were obtained with an indirect method using two forecasting models, one for the overall load and the other for the PV generation. However, the difference with respect to the findings obtained with a unique net load forecasting model was essentially minimal, which may be of great importance for power system distributors or other agents who do not have access to behind-the-meter generation data. A combinatorial machine learning model was employed in (Hu et al., 2023) to forecast short-term power load utilizing a dynamic adjustable weight. Firstly, a mixed machine learning model was developed employing three types of algorithms comprising the upgraded long and short-term neural network, bagging method, and boosting regression algorithm. Secondly, the dynamic error function and the optimal weight optimization technique were applied so as to balance the contradiction between the speed and accuracy of dynamic adjustment. In addition, a penalty term was introduced to increase the algorithm accuracy and the final prediction results. Aiming to reduce the short-term household load prediction error caused by small load scale and differently residential electricity consumption behavior, a novel hybrid forecasting model based on wavelet threshold denoising (WTD), variational mode decomposition (VMD) and bidirectional long short-term memory (BiLSTM) network was proposed in (Wu et al., 2023). In this hybrid model, WTD was used to denoise the raw data firstly. Then trend feature was extracted by VMD. Finally, the trend feature and historical load data were loaded into BiLSTM model for training and testing. In order to fully use the huge information contained in big data, (Chen et al., 2023) provided a new short-term load forecasting method for new-type power system considering many parameters, which relied on Graph Convolutional Network (GCN) and Long Short-Term Memory network (LSTM). Thus, GCN and LSTM were jointly employed to extract the spatial and temporal characteristics of huge data respectively, and eventually the short-term power load forecasting was realized. The paper (Chiu et al., 2023) proposed to develop a hybrid deep learning models and quantile regression loss function for building load prediction. In academic contribution, the study indicated that Convolutional Neural Network (CNN) can extract meaningful information from high uncertainty power load, and Gated Recurrent Unit (GRU) has the benefit in the times-series forecasting. Furthermore, the proposed hybrid framework outperforms the conventional LSTM and GRU neural network. According to (Wang et al., 2023) the different prediction values, the relevant probability distributions of wind power forecast errors are created using a data-driven way. Then, the polynomial chaos expansion surrogate was designed to assist the probabilistic load margin assessment considering wind power forecast uncertainties. The efficiency of the forecast error model was confirmed using the historical data of realistic wind power stations. The results reveal that the probability distributions of forecast errors vary with the level of forecast values. A novel iterative memory-driven load forecast network (IMDLFN) model was proposed in (Yang et al., 2023) to forecast system-level load in the scenario of weight selection problem of BPNN combination variants. It creatively employs memory-driven units to express long-term trend information gathered from the previous time steps and to autonomously balance long-term memory and short-term memory. This properly deals with long-term dependencies and perfectly circumvents weight selection problem of BPNN combination variants.

A short-term power load forecasting (STPLF) model based on the Improved Whale Optimization Algorithm (IWOA) optimized Kernel Extreme Learning Machine (KELM) was suggested in (Han et al., 2023) to address the challenges of high unpredictability and low forecasting accuracy of electricity loads. The KELM model is developed, and the IWOA is utilized to optimize the core and Basic parameters of the KELM to create the IWOA-KELM power load forecasting model. (Al-ja'afreh et al., 2023) Proposed a novel multi-stage framework for PV and load STF that employs feature creation, feature selection, and optimal hyperparameter tuning preprocessing approaches. An enhanced hybrid CNN-LSTM deep learning model architecture was built in the last stage of the suggested framework. It was recognized in (Zhang et al., 2024) that load forecasting plays a vital role for the smart grids and typically, the users have the combinational load prediction demands with diverse time-scales. However, due to numerous internal or external factors, such as inadequate management, the available load data set is limited. Therefore, the problem becomes as the high-precision combinational load forecast at diverse time-scales but with limited reliable datasets. In (Zhang et al., 2024), taking the work at a city power company of China as an example, an empirical study for this topic was considered and then a robust forecasting system was built. Addressing the challenges posed by the non-linear and temporal characteristics of grid load data, (Liu et al., 2024) introduced a novel ultra-short-term power load forecasting model, integrating Convolutional Neural Networks (CNN), Bidirectional Long Short Term Memory networks (BiLSTM), and an Attention mechanism, referred to as the AC-BiLSTM model. This novel technique combines the power of CNN and BiLSTM to extract spatio-temporal aspects of load data, while the Attention mechanism provides appropriate weights to the hidden states of the BiLSTM model, therefore amplifying key historical load sequence data and reducing information loss. The final output of the model is then decided through a completely linked layer. The paper (Bian et al., 2022) created a short-term power load forecasting approach based on accumulated temperature effect and improved Temporal Convolutional Network. Firstly, the common load determining elements in the short-term forecasting process were examined, the data were cleaned and correlated, and a quantitative model of temperature variables that considered the effect of cumulative temperature was established. Secondly, it considered and divided the multi-dimensional influencing factors, the time series characteristic quantity was formed by taking advantage of the Temporal Convolutional Network (TCN), extracting potential relationships between time-series data and non-time-series data, and the optimal load estimation value was obtained. Finally, the output of the TCN was fused with the non-time-series data to generate a new input feature matrix, and the load prediction was completed by the Back Propagation (BP) Neural Network. Considering that the importance of different training samples is varied, a sample weights assignment approach was proposed in (Xiang et al., 2022) to help the STLF to learn the key sample. At first, the sample similarity was assessed considering the properties of distinct input components. Based on this, training samples were selected. Finally, different training samples were allocated with varied sample weights by the designed weights assignment algorithm. With the proposed strategy, the STLF model was able to focus on the relevant samples. (Le Prince et al., 2023) introduced a unique multi-dimensional hierarchical forecasting method built upon structurally-informed machine-learning regressors. A generic formulation of multi-dimensional hierarchies, unifying spatial and temporal dimensions under a common frame was originally defined. Next, a coherency-informed hierarchical learner was designed built around a custom loss function employing optimal reconciliation approaches. The coherency of the created hierarchical forecasts was then secured utilizing similar reconciliation procedures, offering decision-makers a perspective of the future serving aligned decision. A dynamic ensemble approach was presented in (Yang et al., 2023) to forecast residential short-term load accurately.

The fundamental concept was to employ the state-space techniques to dynamically adjust the weight coefficients used to combine the base models. The dynamic ensemble approach was divided into two steps. In the first step, the least square approach was applied to estimate the weight coefficients. In the second step, the particle filter was applied to exclude the estimation mistake in the first stage and change the weight coefficients dynamically to improve the ensemble accuracy. Then, three heterogeneous models (i.e.vector auto regression model, Gaussian process regression model, and the long short-term memory neural network) were applied as the base models and integrated based on the weight coefficients to anticipate the residential load. The numerical experiments were conducted considering two situations in two public datasets to assess the performance of the suggested technique. A multi-energy loads forecasting model based on the artificial intelligence method was proposed in (Ren et al., 2022). Firstly, based on Copula theory, the nonlinear relationship among different load types themselves, as well as other influencing factors affecting the load demand such as ambient temperature etc. were evaluated. Following this, the input factors for load forecasting were selected according to the influence degree. Secondly, based on LSTM (Long Short-Term Memory) neural network model and its upgraded versions (Stack LSTM and Bidirectional LSTM), the multi-type load forecasting model was built.

To further increase the forecast accuracy under the reality of rolling peak load prediction scenarios for the future month, an aggregated hybrid two-phase decomposition with two-layer prediction model architecture (ICEEMDAN-SE-VMD-SSA-XGBoost-MLR) was suggested in (Gao et al., 2022). First, a fully integrated empirical modal decomposition with improved adaptive noise (ICEEMDAN) was used to initially decompose the historical peak load time series, which leads to several eigenmodal functions (IMFs), and by calculating the sample entropy (SE) of each IMF, the IMFs with similar SEs were aggregated and reconstructed into load components representing different time scales. Then, the variational modal decomposition (VMD) was employed to quadratically decompose the less regular parts of the load components to totally diminish their non-smooth properties. Then, extreme gradient boosting (XGBoost) was used to assist in establishing the feature engineering, combining each load component to form the data set required for the prediction model, and using the sparrow search algorithm (SSA) optimized XGBoost with multiple linear regression (MLR) to construct the first layer prediction model for each component, and the prediction results of each component were superimposed to obtain the preliminary prediction results. (Zulfiqar et al., 2022) suggested a hybrid model that incorporates the multivariate empirical modal decomposition (MEMD) and adaptive differential evolution (ADE) method with a support vector machine (SVM). MEMD allows the decomposition of multivariate data to deteriorate with time to efficiently extract the unique information from exogenous variables over multiple time frequencies to ensure high computational efficiency. The ADE algorithm obtains and optimizes the SVM model's appropriate parameters to successfully avoid trapping into local optimum and produces accurate forecasting results. (Chen et al., 2022). Presented an innovative strategy for selectively leveraging building historical data to identify the amount of data that should be used to train the data-driven model. Firstly, the CV(RMSE) curve of each building reflecting the link between training data duration and forecasting accuracy was obtained utilizing Light GBM. Secondly, clustering techniques such as k-means were utilized to identify buildings that were sensitive to the training data length based on CV(RMSE) curves. Finally, the best training data length for day-ahead forecasting was estimated for each building. In (Agga et al., 2022) two deep learning models were created and tested (LSTM, CNN). Both designs went under various different configurations to witness the impact of adjusting the number of hidden layers on the accuracy of the forecasts. The findings demonstrated that the models performed differently when the numbers of layers changed over the different configurations. The methodology of (Shim et al., 2023) entailed splitting the data into weekdays and weekends, conducting separate forecasts for each, selecting characteristics for use in each process, and refining the forecasting model's parameters using hyperparameter tuning. To choose features, the Shapley Additional Explanations approach was used to determine feature importance, and Pearson correlation coefficients were employed to measure linearity between features. This allowed for the selection of input features of high relevance while avoiding those with high linearity with other features. The parameters of the forecasting model were optimized by Grid search, and the ideal combination of the XGBoost forecasting model's Learning Rate, n estimators, and Max depth was assessed. (Hong et al., 2023) provided a novel strategy, based on a hybrid CNNthat was cascaded with a fully-connected network, to analyze week-ahead daily PLF.

The suggested method combines a systematic grid search together with Adaptive Moment Estimation (Adam) optimizer to create the hybrid model. The grid search tunes the network structure and hyperparameters (such as kernel size) of the hybrid CNN while Adam optimizer changes the synaptic weights and parameters (e.g., values of a kernel). Realistic daily peak load data and meteorological (temperature) data in Taiwan were analyzed using this model. Table 1 below shows the merits and demerits of all methods and models reviewed in this paper that can be applied for load forecasting in all energy conversion devices such as converters (A. U. Lawan et al 2013- A. U. Lawan et al 2014)

Table 1a: machine learning methods/models

S/N	Method/Model	Merits	Demerits
1	Gradient Boosting (GBR/XGBoost)	High accuracy, handles non-linearity, resists overfitting	Computationally intensive, requires hyperparameter tuning
2	k-Nearest Neighbors (kNN)	Simple, no training phase, good for local patterns	Poor scalability, sensitive to noise/irrelevant features
3	Support Vector Regression (SVR)	Robust to outliers, effective in high dimensions	Slow for large datasets, kernel selection critical
4	Convolutional Neural Networks (CNN)	Automatically extracts spatial features, good for grid data	Needs large datasets, less intuitive for pure time-series
5	LSTM/GRU	Captures long-term dependencies, ideal for volatile patterns	High computational cost, prone to overfitting
6	Hybrid Models (e.g., TCN-LSTM, CNN-LSTM)	Combines spatial/temporal features, best accuracy	Complex, resource-heavy, hard to debug
7	ForecastNet	Addresses vanishing gradients, adapts to time-varying data	Limited validation, computationally heavy
8	Federated Learning (FedForecast)	Privacy-preserving, uses edge computing	Communication overhead, potential bias
9	Neuroevolutionary Algorithms	Auto-tunes hyperparameters, handles uncertainty	Extremely compute-intensive

Table 1b: Traditional learning methods/models

S/N	Method/Model	Merits	Demerits
1	Time Series (ARMA, Exponential Smoothing)	Simple, interpretable, fast for small data	Assumes linearity/stationarity, poor with volatility
2	Decomposition (STL/EMD/Wavelet)	Separates trends/noise, enhances ML inputs	Manual tuning, risk of signal loss
3	Physics-Based Models	Explainable, works with limited data	Inflexible, requires detailed system specs
4	Fuzzy Logic	Handles imprecise data, rule-based transparency	Subjective rules, scales poorly
5	Bayesian Neural Networks	Quantifies uncertainty, robust to small data	Computationally intensive, complex approximations

Table 1: Merits and demerits of machine learning methods/models (Table 1a) and non-machine learning methods/models reviewed *Note*

1. ML Methods: Excel in accuracy for complex/short-term forecasts but demand data/resources.
2. Non-ML Methods: Transparent and efficient for long-term/linear trends but lack adaptability.
3. Hybrids: Best performance (e.g., decomposition + LSTM) but trade-off is complexity.

Table 2: shows a structured breakdown of the limitations and performance metrics for each electric load forecasting method/model reviewed in this paper.

Table 2: Structured breakdown of the limitations and performance metrics for each electric load forecasting method/model

S/N	Techniques	Models	Limitations	Performance Metrics (Typical Values)
1	Machine Learning (ML) Models	LSTM/GRU	Computationally heavy, slow training, prone to overfitting with small datasets, black-box nature	MAE: 0.8–2.5% RMSE: 1.2–3.0% MAPE: 2–5%
		CNN	Struggles with pure time-series without spatial structure, requires large datasets	MAE: 1.0–3.0% RMSE: 1.5–3.5%
		XGBoost	Less effective for long-term dependencies, hyperparameter tuning critical	MAE: 1.5–4.0% RMSE: 2.0–5.0% R^2: 0.92–0.98
		Hybrid (TCN-LSTM)	Extremely complex, hard to deploy, high resource demands	MAE: 0.7–2.0% RMSE: 1.0–2.8%
		Federated Learning	Communication overhead, potential bias from local data heterogeneity	MAE: 1.2–3.5% RMSE: 1.8–4.0%
		SVR	Scalability issues with large datasets, kernel selection subjective	MAE: 2.0–5.0% RMSE: 2.5–6.0%
		kNN	Poor scalability (high memory usage), sensitive to irrelevant features	MAE: 3.0–7.0% RMSE: 4.0–8.0%
2	Non-Machine Learning (Non-ML) Models	ARIMA/SARIMA	Assumes linearity, fails with volatile data, manual parameter tuning	MAE: 3.0–8.0% RMSE: 4.0–9.0%
		Physics-Based	Inflexible to sudden changes (e.g., weather), requires detailed system specs	MAE: 4.0–10.0% RMSE: 5.0–12.0%
		Fuzzy Logic	Rule design is subjective, poor	MAE: 3.5–8.5% RMSE: 4.5–10.0%

			scalability for high-dimensional data	
		Bayesian Neural Nets	Computationally intensive, complex posterior approximations	MAE: 1.5–3.5% PICP: 85–95% (for 90% interval)
3	Decomposition-Based Methods	EMD/CEEMDAN	Risk of over-decomposition (mode mixing), computationally expensive	MAE: 1.0–2.5% (when combined with LSTM)
		Wavelet Transform	Manual selection of wavelet basis, signal reconstruction artifacts	MAE: 1.2–3.0% (with XGBoost/LSTM)

Critical Observations on Performance

1. Accuracy Trade-offs: Best accuracy was hybrid models (e.g., TCN-LSTM) achieved lowest MAE (0.7–2.0%) but at high complexity while worst accuracy are traditional methods (ARIMA, physics-based) which struggles with MAE >5%.
2. Data Dependencies: ML models (LSTM, XGBoost) require large datasets but fails catastrophically with sparse data while non-ML methods (ARIMA) work with small data but lacks adaptability.
3. Real-World Readiness: There are deployment barriers in Federated learning and edge AI (TinyML) which are promising but lack industry adoption. Only Bayesian NNs and SHAP-enabled models provide uncertainty quantification.

CONCLUSION

The review of 44 academic papers on electric load forecasting (ELF) highlights significant advancements in methodologies, particularly in machine learning (ML), hybrid modeling and conventional methods/models. However, several research gaps and challenges remain, necessitating further exploration to enhance forecasting accuracy, robustness, and real-world applicability. Below is a detailed discussion of unresolved issues in the literature:

1.Data Quality and Heterogeneity: Many models assumed clean, uniform data, but real-world load data is often noisy, sparse, or missing (e.g., smart meter failures, cyberattacks). Few studies addressed data imputation or adversarial robustness (e.g., Moradzadeh et al., 2022, on cyber-secure forecasting).
2. Explainability vs. Performance Trade-off: Deep learning models (LSTM, CNN) outperform traditional methods but act as black boxes, limiting trust in grid operations. Few papers integrated Explainable AI (XAI) techniques (e.g., SHAP, attention maps) for regulatory compliance.
3. Generalization Across Regions and Time: Models trained on one grid (e.g., ISO-NE) often fail to generalize to others (e.g., emerging economies) which leads to lack of transfer learning or meta-learning frameworks for cross-domain adaptation.
4. Real-Time and Edge Deployment: Most research focused on offline batch processing, instead of real-time forecasting for grid stability. Also, lightweight models (e.g., TinyML, quantized GRU) for edge devices are underexplored.
5. Extreme Event Resilience: Models struggle with concept drift (e.g., COVID-19 demand shocks, extreme weather). Only a few studies (e.g., Bayram et al., 2023's DA-LSTM) addressed dynamic adaptation.
6. Renewable Energy Integration: Most work forecasted gross demand, but net load (demand - solar/wind) is critical for modern grids. This shows Limited use of physics-informed ML to model renewable intermittency.

RECOMMENDATIONS

In this paper, several methods/models for ELF have been reviewed, the influence parameters, international typical cases and the trend of frontier research in the frontier of load forecasting are summarized and analyzed, and the recommendation is that an hybrid and Ensemble Models: Develop adaptive hybrid models by combining the strengths of deep learning (LSTM, CNN), decomposition techniques (EMD,

Wavelet), and optimization algorithms (e.g., evolutionary
algorithms). Explore dynamic ensemble learning where model weights adjust in real-time based on forecast errors. Due to the fact that hybrid models (e.g., TCN-LSTM, CNN-GRU) have shown superior accuracy and ensembles (e.g., XGBoost + LSTM) reduce bias and variance.

2. Edge AI and Federated Learning: Deploy lightweight models (e.g., TinyML, Quantized Neural Networks) on smart meters for real-time forecasting then Expand using federated learning (FedForecast) to handle heterogeneous grid data securely. This is to reduce latency and bandwidth usage in smart grids and also ensure data privacy (no raw data sharing).

3. Handling Extreme Events and Concept Drift: Develop adaptive models that detect concept drift (e.g., COVID-19 demand shifts) Using reinforcement learning (RL) for dynamic model retraining because traditional models fail during black swan events (pandemics, extreme weather) and concept drift detection (e.g., DA-LSTM) improves robustness.

4. Integration of External Data Sources: Incorporate weather, economic, and social media data via multimodal learning Using Graph Neural Networks (GNNs) for spatial load forecasting. Load is influenced by multiple external factors (e.g., temperature, GDP, events) and GNNs captures grid topology e.g. GNN-LSTM for regional load forecasting. Hence future research should prioritize: Hybrid models (e.g., decomposition + deep learning), edge computing and federated learning, integration of external data (GNNs, Transformers) and standardized benchmarks for reproducibility. By addressing these gaps, the next generation of load forecasting models will be more accurate, robust, and practical for real-world power systems.

REFERENCES

[1] Agga, F. A., Abbou, S. A., El Houm, Y., & Labbadi, M. (2022). Short-term load forecasting based on CNN and LSTM deep neural networks. *IFAC-PapersOnLine, 55*(12), 777–781.

[2] Al-Ja'afreh, M. A. A., Mokryani, G., & Amjad, B. (2023). An enhanced CNN-LSTM based multi-stage framework for PV and load short-term forecasting: DSO scenarios. *Energy Reports, 10*, 1387–1408.

[3] Aseeri, A. O. (2023). Effective RNN-based forecasting methodology design for improving short-term power load forecasts: Application to large-scale power-grid time series. *Journal of Computational Science, 101984*, 101984.

[4] Bayram, F., Aupke, P., Ahmed, B. S., Kassler, A., Theocharis, A., & Forsman, J. (2023). DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks. *Engineering Applications of Artificial Intelligence, 123*, 106480.

[5] Bian, H., Wang, Q., Xu, G., & Zhao, X. (2022). Research on short-term load forecasting based on accumulated temperature effect and improved temporal convolutional network. *Energy Reports, 8*, 1482–1491.

[6] Chen, H., Zhu, M., Hu, X., Wang, J., Sun, Y., & Yang, J. (2023). Research on short-term load forecasting of new-type power system based on GCN-LSTM considering multiple influencing factors. *Energy Reports, 9*, 1022–1031.

[7] Chen, Y., & Chen, Z. (2022). Short-term load forecasting for multiple buildings: A length sensitivity-based approach. *Energy Reports, 8*, 14274–14288.

[8] Chiu, M.-C., Hsu, H.-W., Chen, K.-S., & Wen, C.-Y. (2023). A hybrid CNN-GRU based probabilistic model for load forecasting from individual household to commercial building. *Energy Reports, 9*, 94–105.

[9] Deng, X., Ye, A., Zhong, J., Xu, D., Yang, W., Song, Z., ... & Chen, X. (2022). Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Reports, 8*, 8661–8674.

[10] Falces, A., Capellan-Villacian, C., Mendoza-Villena, M., Zorzano-Santamaria, P. J., Lara-Santillan, P. M., Garcia-Garrido, E., ... & Zorzano-Alba, E. (2023). Short-term net load forecast in distribution networks with PV penetration behind the meter. *Energy Reports, 9*, 115–122.

[11] Fan, C., Nie, S., Xiao, L., Yi, L., Wu, Y., & Li, G. (2024). A multi-stage ensemble model for power load forecasting based on decomposition, error factors, and multi-objective optimization algorithm. *International Journal of Electrical Power & Energy Systems, 155*, 109620.

[12] Gao, Z., Yin, X., Zhao, F., Meng, H., Hao, Y., & Yu, M. (2022). A two-layer SSA-XGBoost-MLR continuous multi-day peak load forecasting method based on hybrid aggregated two-phase decomposition. *Energy Reports, 8*, 12426–12441.

[13] Hammad, M. A., Jereb, B., Rosi, B., & Dragan, D. (2020). Methods and models for electric load forecasting: A comprehensive review. *Logistics, Supply Chain, Sustainability and Global Challenges, 11*(1), 51–76.

[14] Han, D., Bai, H., Wang, Y., Bu, F., & Zhang, J. (2023). Day-ahead aggregated load forecasting based on household smart meter data. *Energy Reports, 9*, 149–158.

[15] Han, X., Shi, Y., Tong, R., Wang, S., & Zhang, Y. (2023). Research on short-term load forecasting of power system based on IWOA-KELM. *Energy Reports, 9*, 238–246.

[16] Hong, Y. Y., Apolinario, G. F. D., & Cheng, Y. H. (2023). Week-ahead daily peak load forecasting using hybrid convolutional neural network. *IFAC-PapersOnLine, 56*(2), 372–377.

[17] Hu, X., Tang, X., Zhang, Q., Chen, Z., & Zhang, Y. (2023). Combined machine learning forecasting method for short-

term power load based on the dynamic weight adjustment. *Energy Reports, 9*, 866–873.

[18] Jalali, S. M. J., Arora, P., Panigrahi, B. K., Khosravi, A., Nahavandi, S., Osório, G. J., & Catalão, J. P. S. (2022). An advanced deep neuroevolution model for probabilistic load forecasting. *Electric Power Systems Research, 211*, 108351.

[19] Leprince, J., Madsen, H., Møller, J. K., & Zeiler, W. (2023). Hierarchical learning, forecasting coherent spatio-temporal individual and aggregated building loads. *arXiv [cs.LG]*.

[20] Li, X., Wang, Y., Ma, G., Chen, X., Shen, Q., & Yang, B. (2022). Electric load forecasting based on Long-Short-Term-Memory network via simplex optimizer during COVID-19. *Energy Reports,8*,1–12.

[21] Li, Y., Guo, X., Gao, Y., Yuan, B., & Wang, S. (2022). Short-term power load probabilistic interval multi-step forecasting based on ForecastNet. *Energy Reports, 8*, 133–140.

[22] Liu, Y., Dong, Z., Liu, B., Xu, Y., & Ding, Z. (2023). FedForecast: A federated learning framework for short-term probabilistic individual load forecasting in smart grid. *International Journal of Electrical Power & Energy Systems, 152*, 109172.

[23] Liu, F., & Liang, C. (2024). Short-term power load forecasting based on AC-BiLSTM model. *Energy Reports, 11*, 1570–1579.

[24] Mokarram, M. J., Rashiditabar, R., Gitizadeh, M., & Aghaei, J. (2023). Net-load forecasting of renewable energy systems using multi-input LSTM fuzzy and discrete wavelet transform. *Energy, 275*

[25] A. U. Lawan, H. Abbas, J. G Khor (2019), "Enhanced decoupled Current Control with Voltage Compensation for Modular Multilevel Converter (MMC) based STATCOM" International journal of Power Electronics and Drives (IJPEDS) 10 (3), 14-19

[26] A. U. Lawan, Haider A.F. Almurib, Jeen G. Khor (2019), "Modular Multilevel Converter (MMC) based STATCOM with Vector Control and Virtual Impedance Voltage Compensations" International Journal of Power Electronics and Drives (IJPEDS) 10 (4), pp. 51~70

[27] S. Babani, Jazuli, Tonga, A. U. Lawan (2015) "Design and simulation of coupled line coupler with different values of coupling efficiency" International journal of Electrical, Electronics and data communications, Institute of Research and Journals, Switzerland , 3(8), Page 14-17.

[28] A.U. Lawan A.A. Karim I.S Madugu, A.Kunya , G. Shehu (2019) "Modular Multilevel Converter with Harmonics Compensation for State-Feedback Controller" Zaria Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria – Nigeria. 8(2), pp 38-49.

[29] A.U. Lawan A.A. Karim, G. B Shehu, A. Kunya, I.S Madugu (2019)" Modular Multilevel Converter based on Feedforward Disturbance Rejection for harmonic current suppression" Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria – Nigeria. 8(2), pp 96-104.

[30] Ibrahim S. Madugu, B. J. Olufeagba, Yinusa A. Adediran, F. Abdulkadir, A. Abdulkarim James. U. Inaku, A. U. Lawan (2019) "A novel model for solar radiation prediction" TELKOMNIKA,), Institute of Advanced Engineering and Science (IAES), Indonesia, 17(6), pp.3110-3119

[31] Ibrahim S. Madugu, B. J. Olufeagba, Yinusa A. Adediran, A. Abdulkarim, F. Abdulkadir, James. U. Inaku, Olalekan Ogunbiyi, O. Ibrahim2, A. U. Lawan, and Mubarak A. Afolayan (2019), " Arma Model for Prediction of Solar Radiation in Kano-Nigeria " Zaria Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria – 8(1), pp 1-8.

[32] G. S. Shehu, A. B. Kunya, I. H. Shanono, A. U. Lawan "Selective Harmonic Elimination For Cascaded H-Bridge Multilevel Single Source Inverter" Zaria Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria . 8(1), pp 40-49.

[33] A. U. Lawan, Babani, A. Abdulkarim, G. S. Shehu, Y. Jibril, I. S. Madugu (2019), "Power Compensation for Vector-based Current Control of a Modular Multilevel Converter (MMC) based STATCOM" International Journal of Power Electronics and Drives (IJPEDS) 10 (4), pp. 71~85

[34] A. Abdulkarim, SM Abdelkader, DJ Morrow, FJ Alaba, A. U. Lawan (2016), "Effect of weather and the hybrid energy storage on the availability of standalone microgrid" International Journal of Renewable Energy Research (IJRER), Gazi University, Turkey 6 (1), 189-198

[35] Abdullahi Y.M, Magaji N, Sabo I.B and Lawan A. U (2015) "Impact of distriuted Generation Systems on Distribution Network: a case Study of Shiroro-Zamfara Radial Network, Nigeria" Nigerian Journal of Solar Energy, Solar Energy Society of Nigeria, Sokoto Energy Reserve Center, Usman Danfodio University, Sokoto 26 (15), 35-42.

[36] Y. M. Abdullahi, N. Magaji, S. B, A. U. LAWAN (2015) "Load Stability Assessment of a Reduced Shiroro-Zamfara Radial Network in Nigeria "Journal of Engineering Technology, Bayero University 10 (02), 81-88.

[37] N Magaji, MW Mustafa, A Dan-Isa, A.U. Lawan (2013) "Static and dynamic location of variable impedance devices for stability studies" American Journal of Applied Sciences 10 (5), 497 (SCOPUS/Q3).

[38] Ali A. S. Musa, U, Babani S., Muhammad S.G, A.U. Lawan "A Review on Architecture Protocols and Technologies that Underpin Green Migration from 4G to 5G" Bayero Journal of Engineering and Technology" pp 48-57 Vol 15(1) 2020.

[39] A.B. Kunya, G.S. Shehu, U.M. Hassan, A.U. Lawan (2019) "Simultaneous Distribution Network Reconfiguration and Optimal Deployment of Distributed Generation" Journal of Applied Materials and Technology 1(1), pp 46-53 (JAMT, University of Riau, Indonesia,

[40] Dansarki Nanly W, A Abdulkarim , P.U Okorie, Adamu Halilu , A. U.Lawan et al "A Review on Optimal Siting and Sizing of DSTATCOM" Journal of Applied Materials and Technology 1(1), pp 46-53 (JAMT, University of Riau, Indonesia)

[41] Nuraddeen Magaji , Mohd Wazir Bin Mustafa , A. U. Lawan , Alliyu Tukur, Ibrahim Abdullahi and Mohd Marwan "Application of Type 2 Fuzzy for Maximum Power Point Tracker for Photovoltaic System" journal of MDPI-Processes V10(8) 1530, 2022

[42] N. Magaji, Mohd Marwana , A. U lawan, S. B. Ibrahima , MS Gaya. and Nuruddeen salahuddena "Dynamic Modeling of alkaline electrolyzer Based on Matlab Simulink approach" International Review on Energy Conversion (IRECO.), Vol. 14. June 2022

[43] A. U Lawan, HM Abbas (2015) "Level shifted PWMs comparison for a 5-level modular multilevel converter MMC) topology inverter" IEEE CSUDET Malaysia 1-6.

[44] A. U Lawan, M Mustapha, I Abubakar (2015) "Reactive current control of STATCOM based MMC inverter for wind turbines connected to grid "IEEE Conference (SCOReD) Malaysia, 26-31

[45] A. U. Lawan, N Magaji, H Musa (2013) " A STATCOM controller for small signal stability using polynomial algorithms in a horizontal axis wind Farm power system" ENERGYTECH, IEEE U.S.A, 1-