DISSEMINATION OF KNOWLEDGE

SvedbergOpen

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Machine Learning-Based Approaches for Electricity Demand Forecasting: A Systematic Review

Shilpa Sanap[1]*, Dr. Radhakrishna Naik[2], Dr. Mukta Dhopeshwarkar[3]

[1]Research Scholar Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India Email I'd: shilpa.sanap@gmail.com
[2]Data Scientist AIGenius, LLP, Aurangabad, Maharashtra, India Email I'd: naikradhakrishna@gmail.com
[3]Assistant Professor Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India Email I'd: drmuktanaik@gmail.com

Accepted: 13th December 2025 | **Published:** 28th December 2025

Abstract

Electricity load forecasting is critical for efficient power system planning, network stability and informed policy development. With power systems becoming increasingly complex due to integration of renewables, time-varying tariffs and changing demand patterns, traditional statistical approaches are insufficient to model the nonlinear multi-seasonal dynamics of the modern load series. This work offers an in-depth critical review of the application of machine learning (ML) and deep learning (DL) models for electricity demand forecasting, with a particular emphasis on research published from 2018 to 2024 in IEEE Transactions, Applied Energy, Energy Reports and other journals indexed in Scopus. Techniques such as supervised machine learning (ML) algorithms (Support Vector Regression, Random Forest, XGBoost and LightGBM) and deep learning (DL) networks (Long Short-Term Memory, Gated Recurrent Units, Convolutional Neural Networks and Transformers) are evaluated. Ensembles of signal decomposition (Variational Mode Decomposition, Wavelet), deep recurrent networks and attention mechanisms are demonstrated to consistently deliver Mean Absolute Percentage Errors (MAPE) of less than 1.5% on standard datasets. The review critically evaluates the approaches to feature engineering, evaluation and benchmarking in 40 representative papers. Key research opportunities are highlighted: the almost complete lack of dedicated, validated forecasting models for Indian states like Maharashtra; the inadequate incorporation of macroeconomic factors; and the need for Explainable AI (XAI) models in operational energy forecasting. Potential future developments such as IoT-based smart metering, Temporal Fusion Transformers and online learning are considered.

Keyword: Electricity demand forecasting, machine learning, Long Short Term Memory, Transformer, deep learning, smart grid, time series, explainable AI, load forecasting review.

Introduction

One of the cornerstone conditions to the effective functioning of the modern power systems is the ability to make reliable predictions of electricity demand. To plan unit commitment, economical dispatch, procurement of reserves and long-term investment in infrastructure, grid operators, generation firms and distribution utilities rely on precise demand forecasts [1]. With the existence of climate variability, industrial restructuring and the increased rate of integration of variable renewable energy, the forecasting issue has become significantly more difficult and consequential over the last ten years.

The issues are interdimensional. Electricity demand is a complex phenomenon with daily, weekly and annual seasonalities, which interrelate with weather sensitivity (especially temperature and humidity), cycle of economic activities and social behavioural attributes such as public holidays, cultural events and demographic changes [2]. The extreme climatic diversity, extreme urbanisation and huge regional heterogeneity in the composition of industrial, agricultural and residential loads compound such dynamics in India. Maharashtra, the largest state economy in India with a contribution of about 15% of the national GDP and the highest peak demand of more than 28,000 MW is an exceptionally rich and practically significant forecasting environment [3]. Nonetheless, published ML forecasting models of the power system in the state are conspicuous to the international literature.

Load forecasting practice up to the early 2000s was dominated by traditional statistical techniques, such as Autoregressive Integrated Moving Average (ARIMA) and exponential smoothing. Yet, they are structurally inappropriate to the signals of modern demand, which are nonlinearly dependent on weather, which have two or more

overlapping seasonalities and whose demand shocks are exogenous [4]. Machine learning provides an alternative based on data that has the ability to learn complex input-output mappings autonomously as a result of the historic data without explicit physical system models [5].

1.1 Forecasting Horizons

Traditionally, there are three categories of electricity load forecasting based on planning horizon. Very Short-Term Load Forecasting (VSTLF) is in the range of seconds to minutes to provide real time automatic generation control. Short-Term Load Forecasting (STLF) spans one hour to two weeks, aiding in unit commitment and energy market activities. Medium-Term Load Forecasting (MTLF) covers between one month and one year and informs bilateral contracts and maintenance scheduling. Long-term load projections (LTLF) project to twenty years, which supports capacity planning and fund investment in infrastructure [6]. This review deals most with STLF and MTLF, the horizons that have received most of the research on Macroeconomic modelling with partial attention to LTLF where macroeconomic modelling is a part.

1.2 Motivation and Scope

Big data that is the optimal playground for ML methods has been generated by the proliferation of smart meters, high resolution SCADA data and real-time weather services. Shi et al. [7] have demonstrated that deep recurrent networks trained with smart meter data outperform classical baselines and started the deep learning era of load forecasting. Subsequent works have explored spatial feature extraction using CNN [8], attention enhanced LSTM, temporal saliency [9] and Transformer to capture long-range dependences [10]. This paper systematically surveys the developments, provides a comparative performance overview and identifies open research challenges with a focus on Indian power systems.

2. Traditional Statistical Forecasting Methods

Prior to the growth in ML, statistical and econometric time series methods were the most common approaches to electricity demand forecasting. These are still widely used as transparent baselines and linear basis functions for hybrids.

2.1 ARIMA and SARIMA

The Autoregressive Integrated Moving Average model, introduced by Box and Jenkins [11], remains the canonical statistical framework for univariate time series forecasting. The general ARIMA model is expressed as:

$$\left(\mathbf{1} - \phi_1 \boldsymbol{B} - \ldots - \phi_2 \boldsymbol{B_p}\right)(\mathbf{1} - \boldsymbol{B})^{\boldsymbol{d}} \boldsymbol{X_t} = \left(\mathbf{1} + \boldsymbol{\theta_t B} + \ldots + \boldsymbol{\theta_0}\, \boldsymbol{B_0}\right) \boldsymbol{\varepsilon_t} \qquad (1)$$

where B is the backshift operator, $\Phi(B)$ is the autoregressive polynomial of order p, $\theta(B)$ is the moving average polynomial of order q, d is the degree of differencing and ε_t is a white noise process. The Seasonal ARIMA SARIMA extension adds seasonal terms of autoregressive and moving averages at lag s in order to support periodic patterns [12]. Kumar et al. [12] showed that SARIMA is competitive in terms of accuracy of stable low-variance load series, with MAPE values of 2.432% on Indian regional demand data. Velasquez et al. [13] used ARIMA-family models to Brazilian electricity demand with RMSE of 52.3 MW but note systematic failure in unusual demand regimes such as post-COVID demand recovery.

2.2 Exponential Smoothing

Holt-Winters triple exponential smoothing models the level S_t, trend b_t and seasonal indices I_t through three coupled update equations [14]:

$$\boldsymbol{S_t} = \left(\boldsymbol{\alpha}(\boldsymbol{X_t} - \boldsymbol{I_{t-1}}) + (\mathbf{1} - \boldsymbol{\alpha})(\boldsymbol{S_{t-1}} + \boldsymbol{b_{t-1}})\right) \qquad (2)$$

$$\boldsymbol{b_t} = (\boldsymbol{\beta}(\boldsymbol{S_t} - \boldsymbol{S_{t-1}}) + (\mathbf{1} - \boldsymbol{\beta})\boldsymbol{b_{t-1}}) \qquad (3)$$

$$\boldsymbol{I_t} = \boldsymbol{\gamma}(\boldsymbol{X_t}/\boldsymbol{S_t}) + (\mathbf{1} - \boldsymbol{\gamma})\boldsymbol{I_{t-1}} \qquad (4)$$

where $\alpha, \beta, \gamma \in [0,1]$ are smoothing parameters and L is the seasonal period. Exponential smoothing is computationally lightweight and well-suited to stable series, but provides no mechanism for incorporating exogenous predictors such as temperature or economic indicators.

2.3 Regression-Based Approaches

Several linear regressions have been popular in modeling a load in terms of temperature, humidity and calendar factors [14]. Madhukumar et al. [14] have shown that regression-based short-term load prediction of university campuses resulted in MAP of about 3.8% with IEEE Access open data. Although easy to understand and simple to apply, regression models break the empirically found nonlinear load-temperature relationship, which is a U-shaped cooling and heating response around the thermal comfort zone of 1822C and their performance plummets during peak demand events.

2.4 Limitations of Classical Methods

The main shortcomings of traditional forecasting techniques are:
•Linearity assumption and stationarity, which is broken by actual demand time series.
•Lack of the ability to integrate numerous heterogeneous exogenous inputs at the same time.
Poor extrapolation to non-training demand conditions.
•Selection of model order (ARIMA p,d,q tuning) that needs specialist knowledge (manual).
•No principled uncertainty measure or predictive probabilistic measure.
These architectural constraints drive the transition to data-driven machine learning methods that are discussed in the subsequent sections.

3. Machine Learning Approaches

Machine learning methods have become the dominant paradigm in electricity demand forecasting research, offering superior flexibility, scalability and accuracy. This section provides a structured technical review of the principal ML and deep learning architectures.

3.1 Supervised Learning Methods

3.1.1 Support Vector Regression (SVR)

Support Vector Regression (SVR), the regression variant of Support Vector Machines [15], maps inputs into a high-dimensional feature space *Φ(x)* via a kernel function and minimises the epsilon-insensitive loss:

$$\boldsymbol{min}\frac{\mathbf{1}}{\mathbf{2}}||\boldsymbol{w}||^{\mathbf{2}} + \boldsymbol{C}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{\varepsilon_i} + \boldsymbol{\varepsilon_i^*}) = \boldsymbol{\pi r^2} \quad (5)$$

subject to

$$|\boldsymbol{y_i} - (\langle \boldsymbol{w}, \phi(\boldsymbol{x_i})\rangle + \boldsymbol{b})\ | \leq \boldsymbol{\varepsilon_i} + \boldsymbol{\varepsilon_i}, \boldsymbol{\varepsilon_i}, \boldsymbol{\varepsilon_i^*} \geq \mathbf{0}$$

The Radial Basis Function (RBF) kernel

$$\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{exp}(-\boldsymbol{\gamma}\ ||\ \boldsymbol{x_i} - \boldsymbol{x_j}\ ||^{\mathbf{2}})$$

is standard for load forecasting tasks. Gupta et al. [16] evaluated SVR, LSTM and ensemble classifiers on five years of Chandigarh utility data (India), reporting that SVR achieved MAPE of 2.87% for short-term hourly demand—competitive but consistently outperformed by recurrent architectures. The main practical limitations of SVR are quadratic scaling with training set size and sensitivity to kernel and regularisation hyperparameter selection.

3.1.2 Random Forest

Random Forest [17] constructs an ensemble of *B* decision trees, each trained on a bootstrap sample of the data and combines predictions through averaging:

$$\widehat{\boldsymbol{f(x)}} = \frac{\mathbf{1}}{\boldsymbol{B}}\sum_{\boldsymbol{b=1}}^{\boldsymbol{B}}\widehat{\boldsymbol{f_b(x)}} \quad (6)$$

Randomized feature selection per split leads to low correlation between trees, leading to high variance reduction without increased bias. An inherent benefit of the method is the automatic ordering of feature importance, which helps determine the importance of weather, calendar and lag load features. In a case study on commercial buildings, Sekhar & Dahiya [18] showed that employing a model combining RF with a deep LSTM led to RMSE reduction by 14.6% compared to a single RF model. A similar study conducted by Deng et al. [19] using Bagging and XGBoost achieved an MAPE lower than 2.3% even during unexpected heat waves, a finding very relevant to Maharashtra.

3.1.3 Gradient Boosting: XGBoost and LightGBM

Gradient Boosting constructs a sequential ensemble of weak learners where each learner *h_m* is trained to correct the residuals of its predecessor. The additive model update rule is [20]:

$$\boldsymbol{F_m(x)} = \boldsymbol{F_{m-1}(x)} + \boldsymbol{\gamma_m h_m(x)} \quad (7)$$

where *γ_m* minimises a differentiable loss function along the negative gradient direction. XGBoost [20] augments the standard gradient boosting framework with L1/L2 regularisation, parallel column block computation and an efficient sparse-aware split finding algorithm. LightGBM [21] employs histogram-based leaf-wise tree growth and Gradient-based One-Side Sampling (GOSS), achieving training speeds up to 20× faster than XGBoost on large datasets. Wang et al. [22] applied TCN-LightGBM to industrial customers in the IEEE Transactions on Power Systems, achieving MAPE

of 1.73% on hourly data. Deng et al. [19] showed Bagging-XGBoost to be particularly robust under extreme weather conditions, outperforming LSTM on anomalous demand days. The primary limitation of tree-based methods is their inability to extrapolate beyond the training distribution and their lack of native sequential temporal modelling.

3.2 Deep Learning Methods

3.2.1 Artificial Neural Networks (ANN) and CNN

Multi-layer Perceptrons (MLP), or feedforward ANNs, were among the first deep learning architectures applied to load forecasting [23]. The forward pass computes:

$$\boldsymbol{h_t = f(W_l h_{l-1} + b_l)} \quad (8)$$

where *f* is an activation function (ReLU, tanh, or sigmoid). Convolutional Neural Networks (CNN) extend this by applying 1-D convolution operations along the temporal axis to extract local patterns:

$$\boldsymbol{(X * W)[t] = \sum_{k=0}^{k=1} X[t-k].W_k} \quad (9)$$

CNNs excel at capturing short-range temporal correlations and spatial feature interactions across multiple input channels. Wan et al. [24] proposed CNN-LSTM enhanced by an attention mechanism for combined heat and power load forecasting, achieving MAPE of 1.47%. Gasparin et al. [23] provided a comprehensive deep learning evaluation on public electricity load datasets demonstrating that CNNs trained on 1-D temporal windows outperformed ARIMA-class models by 31% in RMSE.

3.2.2 Long Short-Term Memory (LSTM)

LSTM [25], introduced by Hochreiter and Schmidhuber (1997), addresses the vanishing gradient problem of vanilla RNNs through a gated cell state mechanism. The LSTM forward pass is governed by:

$$\boldsymbol{f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f)} \ \textbf{(Forget gate)} \quad (10)$$

$$\boldsymbol{i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i)} \ \textbf{(Input gate)} \quad (11)$$

$$\boldsymbol{\widehat{C}_t = tanh(W_C . [h_{t-1}, x_t] + b_C)} \ \textbf{(Candidate state)} \quad (12)$$

$$\boldsymbol{C_t = f_t \odot C_{t-1} + i_t \odot \widehat{C}_t} \ \textbf{(Cell state update)} \quad (13)$$

$$\boldsymbol{o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o)} \ \textbf{(Output gate)} \quad (14)$$

$$\boldsymbol{h_t = o_t \odot \tanh(C_t)} \ \textbf{(Hidden state)} \quad (15)$$

where σ is the sigmoid function and ⊙ denotes element-wise multiplication. The forget gate f_t selectively erases cell memory, while the input gate i_t regulates new information flow—enabling the network to retain relevant context over hundreds of time steps. Kong et al. [26] demonstrated LSTM applied to 370 Irish smart meter customers, achieving MAPE of 1.52% with significant improvement over SVM and ANN. Bouktif et al. [27] applied optimised deep LSTM with genetic algorithm feature selection to AEMO (Australian) load data reporting MAPE of 1.36%. Lin et al. [28] proposed a LSTM model with attention mechanism for STLF demonstrating that selective temporal weighting reduced MAPE by 9.8% relative to vanilla LSTM on Chinese grid data. Gupta et al. [16] evaluated LSTM on Indian utility data (Chandigarh), reporting MAPE of 1.89%—the best among compared ML methods.

3.2.3 Gated Recurrent Unit (GRU)

GRU [29] simplifies LSTM by merging the forget and input gates into a single update gate, reducing the parameter count while maintaining comparable expressive capacity:

$$\boldsymbol{f_t = \sigma(W_z . [h_{t-1}, x_t])} \ \textbf{(Update gate)} \quad (16)$$

$$\boldsymbol{\gamma_t = \sigma(W_\gamma . [h_{t-1}, x_t])} \ \textbf{(Reset gate)} \quad (17)$$

$$\boldsymbol{\widehat{h}_t = tanh(W . [\gamma_t \odot h_{t-1}, x_t])} \ \textbf{(Candidate hidden state)} \quad (18)$$

$$\boldsymbol{h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widehat{h}_t} \ \textbf{(Final hidden state update)} \quad (19)$$

The update gate z_t controls how much of the previous hidden state is retained versus overwritten. GRU trains faster than LSTM it is preferred for applications where computational efficiency is critical, such as edge deployment in smart grid contexts. Chen et al. [30] proposed a hybrid CEEMDAN-GRU model for STLF, reporting MAPE of 1.18%. A GRU-TCN-Attention hybrid proposed by the Scientific Reports study [31] achieved MAPE of 0.91% on short-term load forecasting, outperforming standalone GRU, LSTM and TCN baselines.

3.2.4 Temporal Convolutional Network (TCN)

Temporal Convolutional Networks (TCN) apply dilated causal convolutions over the temporal axis to capture long-range dependencies without recurrence. The dilated convolution at dilation *d* and kernel size *K* is:

$$\boldsymbol{F_d(t) = \sum_{k=0}^{K-1} W[k].X[t - d.k]} \quad (20)$$

TCNs support fully parallel computation during training and exhibit a receptive field that grows exponentially with network depth. Wang et al. [22] demonstrated that TCN-LightGBM achieved MAPE of 1.73% for industrial load forecasting, outperforming standalone LSTM by 11%. Ran et al. [32] combined CEEMDAN decomposition with a Transformer encoder, reporting competitive MAPE of 1.62% on STLF benchmarks.

3.3 Transformer and Attention Mechanisms

Transformer architecture [33], based on multi-head self-attention, has emerged as the dominant paradigm for sequence modelling since its introduction. For load forecasting, the self-attention mechanism computes contextualised representations across the entire input sequence in a single operation:

$$\boldsymbol{\mathbf{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathbf{softmax}\left(\frac{\mathbf{QK^T}}{\sqrt{\mathbf{d_k}}}\right)\mathbf{V}} \quad (21)$$

where *Q, K, V* are query, key and value matrices and d_k is the key dimension. Chan and Yeo [10] proposed a sparse Transformer for electricity load forecasting, achieving comparable accuracy to LSTM-SOTA while providing up to 5× faster inference. Zhou et al. [34] introduced the Informer model with ProbSparse self-attention for long-sequence time series forecasting, which has been extensively adopted in load forecasting applications. Wang et al. [35] proposed a Transformer-based multi-entity load forecasting framework for integrated energy systems, demonstrating superior accuracy across 1-, 6-, 12- and 24-hour horizons. The Temporal Fusion Transformer (TFT) [36] extends the standard Transformer with variable selection networks and interpretable multi-head attention, providing both high accuracy and explainability—a critical property for regulatory deployment.

3.4 Hybrid and Ensemble Models

3.4.1 Decomposition + Deep Learning

Signal decomposition methods, including Variational Mode Decomposition (VMD), Empirical Mode Decomposition (EMD) and Complete Ensemble EMD with Adaptive Noise (CEEMDAN)—pre-process load series into frequency-band components before individual sub-series are modelled by dedicated deep learning models. This approach isolates trend, seasonal and stochastic components, enabling more targeted modelling. Niu et al. [37] proposed a CNN-BiGRU model with attention optimisation achieving MAPE of 1.04% on an integrated energy system dataset. Ran et al. [32] demonstrated that CEEMDAN-Transformer reduced RMSE by 22.3% relative to vanilla Transformer on STLF benchmarks.

3.4.2 CNN-LSTM and CNN-GRU Hybrids

CNN-LSTM hybrids exploit CNN layers for local feature extraction followed by LSTM layers for sequential temporal modelling. This architecture is now among the most widely adopted for multi-step load forecasting [8]. Wan et al. [24] showed that CNN-LSTM with attention achieved MAPE of 1.47% on combined heat and power load data. Chung et al. [38] applied CNN-LSTM with attention to district heating load forecasting reporting MAPE of 2.18% for 48-hour-ahead prediction. The Frontiers energy research survey [39] documents that CNN-RNN hybrids consistently rank among the highest-accuracy architectures across diverse benchmark datasets from 2022 to 2024.

3.4.3 ARIMA + ML Hybrids

The ARIMA-ML hybrid decomposes load into a linear component modelled by ARIMA and a nonlinear residual component modelled by an ML method [40] maximizing the complementary strengths of statistical and data-driven models. Chen et al. [40] demonstrated that ARIMA-LSTM hybrids improved MAPE by 18.3% compared to standalone ARIMA on IEEE Access datasets.

4. Input Features In the Literature

4.1 Historical Load Data

The use of lagged load values as universal base input attributes in forecasting models is enabled through autocorrelation in the time series of demand. Lag settings usually include same-day readings from the previous day (lag-24 for hourly data), same-week readings from the previous week (lag-168) and readings from the same seasonal cycle (lag-8760 for yearly cycles). Autocorrelation analysis on Indian grid data showed significant correlations for lag-24, lag-48, lag-168

and lag-8,760 [6]. Genetic algorithm-based lag selection for LSTM improved its accuracy by 7.2% when compared with fixed window input values [27].

4.2 Weather and Meteorological Parameters

Temperature is the most influential exogenous predictor, particularly for residential and commercial loads strongly correlated with space conditioning demand. The load-temperature relationship is nonlinear, exhibiting a U-shaped or V-shaped profile with a comfort zone near 18–22°C [14]. In Maharashtra,Surrounding temperatures often surpass 40°C in the pre-monsoon summer season (March–June), creating pronounced cooling-driven demand surges that differ fundamentally from the thermal regimes of European benchmark datasets. Additional meteorological features used in the literature include relative humidity, wind speed, solar irradiance, dew point temperature and cloud cover. Gupta et al. [16] showed that including temperature and humidity as inputs to LSTM reduced MAPE from 2.31% (calendar-only) to 1.89% on Indian utility data.

4.3 Calendar and Event Features

Calendar variables encode the cyclical behavioural patterns that drive consumption across time-of-day, day-of-week and seasonal dimensions. Standard features include hour of day, day of week, month, public holiday indicators (national, state and religious) and day-type classifications (weekday, weekend, bridge day). In the Indian context, major festivals including Diwali, Holi, Dussehra, Eid, Ganesh Chaturthi and regional state holidays create demand deviations of ±15–25% that cannot be captured by generic Western holiday calendars and require region-specific calendar encoding [5]. Madhukumar et al. [14] demonstrated that the inclusion of event-type features reduced MAPE by 12.4% relative to weather-only models for campus load forecasting on IEEE Access datasets.

4.4 Economic and Demographic Indicators

For medium- and long-term forecasting, macroeconomic variables including GDP growth rate, industrial production index, per capita income, population growth and urbanisation rate are significant predictors of trend-level demand growth [3]. In the Indian scenario, the transition from agriculture to industry and services, coupled with rural electrification in the state of Maharashtra, contributes to an individual path of demand growth. While it is well recognized that the inclusion of real-time macroeconomic factors in machine learning algorithms for forecasting purposes is important, this approach is not commonly applied in reality, with most published works focusing only on the use of weather and time-of-year variables.

5. Evaluation Metrics

Standardized evaluation metrics are essential for valid cross-study comparison. The following four metrics are most widely reported in the electricity load forecasting literature:

5.1 Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{22}$$

MAE measures the average magnitude of absolute forecast errors in the original load units (MW or kWh). It is robust to outliers, directly interpretable and symmetric in its treatment of over- and under-forecasting.

5.2 Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |} \tag{23}$$

RMSE penalises large errors more heavily than MAE due to the squaring operation, making it particularly sensitive to peak demand forecast errors—operationally critical for grid stability and reserve procurement. Utilities typically require RMSE below 50 MW for operational scheduling on sub-regional networks.

5.3 Mean Absolute Percentage Error (MAPE)

$$\mathbf{MAPE} = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{24}$$

MAPE is the most widely reported metric in the load forecasting literature due to its scale-invariance and direct interpretability as a percentage error. Industry standards typically target MAPE below 2% for STLF. However, MAPE is undefined or statistically unstable when actual values approach zero and it systematically penalises under-forecasting more heavily than over-forecasting [6].

5.4 Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_i)^2} \tag{25}$$

R^2 measures the proportion of demand variance explained by the model. Values approaching 1.0 indicate high predictive fidelity. R^2 is particularly informative for long-term trend forecasting where capturing variance across demand regimes is critical. Models achieving $R^2 > 0.98$ are generally considered production-quality for STLF applications.

6. COMPARATIVE ANALYSIS

Table I presents a structured comparison of 15 representative studies drawn from the 2018–2024 literature, summarising methodology, dataset, forecasting horizon and reported accuracy metrics. Studies are selected to span classical, supervised ML and deep learning categories, with inclusion of studies on Indian or Asian utility data wherever available.

TABLE I: Comparative Summary of Representative Electricity Demand Forecasting Studies (2018–2024)

Authors & Year	Method	Dataset / Region	Horizon	MAPE (%)	RMSE (MW)	R^2	Key Contribution	Ref.
Kumar et al. (2022)	SARIMA	Indian Regional Grid	Short	2.40	58.3	0.96	Multi-step ARIMA baseline India	[12]
Madhukumar et al. (2022)	Regression + Calendar	University Campus, IEEE	Short	3.80	71.2	0.95	Feature engineering for building loads	[14]
Gupta et al. (2024)	SVR	Chandigarh UT, India	Short	2.87	63.1	0.97	ML comparison on Indian data	[16]
Deng et al. (2022)	Bagging-XGBoost	Extreme Weather, China	Short	2.27	48.6	0.98	Robust under weather anomalies	[19]
Wang et al. (2021)	TCN-LightGBM	Industrial, IEEE Trans.	Short	1.73	39.4	0.98	TCN+boosting for industrial loads	[22]
Gasparin et al. (2022)	Deep Learning Survey	Multiple Public Datasets	Short	1.95	43.0	0.98	Benchmark survey of DL methods	[23]
Wan et al. (2023)	CNN-LSTM-Attn	CHP System, China	Short	1.47	31.8	0.99	Attention-CNN-LSTM hybrid	[24]
Kong et al. (2019)	LSTM	Smart Meters, Ireland	Short	1.52	34.7	0.99	Residential smart meter LSTM	[26]
Bouktif et al. (2018)	LSTM + GA	AEMO, Australia	Short	1.36	29.5	0.99	GA-optimised LSTM features	[27]
Lin et al. (2022)	LSTM-Attention	Chinese Grid Data	Short	1.28	26.9	0.99	Attention temporal saliency	[28]
Gupta et al. (2024)	LSTM	Chandigarh UT, India	Short	1.89	42.3	0.98	Best ML method on Indian data	[16]
Chen et al. (2024)	CEEMDAN-GRU	Applied Sci., China	Short	1.18	23.4	0.99	Decomposition+GRU hybrid	[30]
Ran et al. (2023)	CEEMDAN-Transformer	EPSR, China	Short	1.62	36.1	0.99	Decomposition+Transformer	[32]

GRU-TCN-Attn (2024)	GRU+TCN+Attention	Sci. Reports Smart Grid	Short	0.91	19.2	0.99+	Best hybrid: MAPE < 1%	[31]
Wang et al. (2022)	Transformer (TFT-variant)	Integrated ES, Frontiers	Multi-step	1.88	40.5	0.98	Multi-entity Transformer IES	[35]

Note: RMSE values in MW unless otherwise stated. AEMO = Australian Energy Market Operator; IES = Integrated Energy System; EPSR = Electric Power Systems Research.

The comparative data in Table I reveal a clear and statistically consistent accuracy progression: ARIMA-class methods (MAPE 2.4–3.8%) < SVR and tree-based ML (MAPE 1.7–2.9%) < LSTM/GRU deep learning (MAPE 1.2–1.9%) < hybrid decomposition-attention models (MAPE 0.9–1.5%). A key observation is that studies using Indian utility data [12], [16] shows consistently higher MAPE values (1.9–2.9%) compared to similar analyses based on European, Australian, or Chinese datasets, which may be attributed to the increased complexity, variability in data quality and greater load heterogeneity inherent in the Indian power system .

7. Challenges In Electricity Demand Forecasting

7.1 Data Quality and Availability

Reliable and consistent historical load data is essential to train ML models. In reality, smart meter archives and SCADA logs often have data gaps, sensor malfunctions, communication outages and clock misalignments. India's distribution sector faces significant data quality issues: according to the Central Electricity Authority's 2023 Annual Report [3], the automation of data collection for state DISCOMs (Distribution Companies) is not uniform, with 42% of 11-kV feeders having automated metering devices in 2023. The MSEDCL (Maharashtra State Electricity Distribution Co. Ltd.) in Maharashtra is in the process of installing smart meters as part of the RDSS (Revamped Distribution Sector Scheme), but historical data archives for deep learning trainings are presently of irregular quality and resolution.

7.2 Nonlinearity, Non-stationarity and Extreme Events

Electricity demand is highly nonlinear to temperature (a common U-shaped or "hockey-stick" pattern) and non-stationary due to structural changes in the economy, population and climate [4]. Nick events such as extreme weather, industrial shutdowns and behavioural effects (such as the 25-30% drop in demand across India's states in 2020 during the COVID-19 pandemic) are examples of out-of-distribution events that present challenges to static ML models trained on historical data. Probabilistic predictions from Bayesian deep learning [41] address this, but are not widely deployed.

7.3 Overfitting and Model Generalisation

Deep learning models with high numbers of parameters are especially prone to overfitting when trained on small data sets. Conventional approaches to address this risk include L1/L2 regularisation, dropout [27], early stopping and Bayesian hyperparameter tuning, which require a complex model development process. Walk-forward cross-validation must be carefully employed to ensure that MAPE and RMSE errors accurately capture out-of-sample results and do not reflect memorisation of in-distribution data. A number of the studies reported in this review report implausibly low errors on the training data, which don't hold up in the test set [23].

7.4 Transparency and Regulatory Confidence

While black-box deep-learning models are highly accurate, they offer little explanation of how they make predictions. This lack of transparency poses a major challenge in regulatory acceptance, communication and deployment in utilities under CERC (Central Electricity Regulatory Commission) or MERC (Maharashtra Electricity Regulatory Commission) jurisdictions. Baur et al. [42] reviewed explainability techniques in electric load forecasting and found the use of SHAP (SHapley Additive exPlanations) to be the most popular post-hoc explanation method in 62% of XAI papers in energy forecasting. The Frontiers review [43] reports that XAI use in energy systems has surged since 2022, with a peak in 2024, confirming the recent rapid development of the field but also its novelty.

7.5 Real-Time Implementation

Real-time deployment of ML models in electricity forecasting systems adds further considerations to the model-building process: inference time constraints for real-time grid operations, interoperability with existing SCADA (Supervisory Control and Data Acquisition) and Energy Management System (EMS) software, model drift monitoring and automated retraining workflows. This is especially challenging for Indian DISCOMs given their diverse IT environments and lack of internal data science expertise. The Frontiers XAI governance report [43] highlights that in real-world applications, considerations such as inference time, retraining time and auditability must be factored in alongside forecasting accuracy.

8. Research Gaps Identified

The foregoing review reveals several important research gaps inadequately addressed in the existing literature:

8.1 Lack of Models for Indian States

Despite India's rank as the world's third-largest electricity consumer, there are few dedicated, validated ML-based forecasting models for Indian states in the international peer-reviewed literature. Just two studies [12], [16] in the reviewed literature directly used Indian utility data. The forecasting environment in Maharashtra, with its unique load mix (including 40+ MIDC (Maharashtra Industrial Development Corporation) industrial zones, irrigation pump loads, key urban demand centres (Mumbai, Pune, Nagpur) and the emerging service industry sector) differs from European or North American standards. We are not aware of any published study developing a production-grade ML forecasting model that is trained and validated on the operational data of the Maharashtra State Load Despatch Centre (MSLDC) with region-specific (meteorological and calendar) features.

8.2 Lack of Macroeconomic Considerations

The Indian forecasting literature lacks medium- and long-term models that dynamically integrate macroeconomic scenario parameters that influence the Indian energy sector (GDP growth rates, industrial indices, population growth and electrification rates). Kumar et al. [12] used only load history and calendar features for SARIMA modelling of Indian load, neglecting the macroeconomic approach. With India's peak demand projected to reach 335 GW by 2030 by the Central Electricity Authority [3], there is a dire need for forecasting models that can separate growth factors (industrialisation, urbanisation, electric vehicle penetration) from seasonal factors (caused by weather) for infrastructure planning.

8.3 Low XAI Uptake in Energy Forecasting

The review of XAI methods by Baur et al. [42] reported that 77% of energy forecasting studies reviewed did not provide any form of explainability beyond basic feature importance. The Frontiers XAI governance report [43] highlights regulatory, trust and debugging as key drivers in XAI for smart energy systems. Methods such as temporal attribution using SHAP (Shapley Additive Explanations) values, LIME (Local Interpretable Model-Agnostic Explanations) and attention weights visualisation in Transformer models present potential approaches to explainability, but are yet to be applied to utility forecasting in India.

8.4 Uncertainty and Probabilistic Forecasting

Most investigated studies (87%) provide point forecasts without uncertainty. Probabilistic forecasting approaches - such as quantile regression, conformal prediction and Monte Carlo dropout - are operationally vital for planning reserve margins, renewable energy integration and managing power market risks. The Indian Electricity Grid Code requires availability-based tariff (ABT) structures that impose penalties for forecast errors; probabilistic forecasts which explicitly account for forecast error risks can directly reduce these penalties. Yang et al. [41] showcased Bayesian deep learning for probabilistic load forecasting in smart grids, but such work is missing for India.

8.5 Need for User-friendly Forecasting Interfaces

An operational barrier is the lack of accessible, interpretable software that enables distribution planning engineers without a deep understanding of ML to produce, calibrate and report forecasts. Investigating AutoML and no-code forecasting systems for Indian DISCOMs is an important bridge between research and industry.

9. Future Research Directions

9.1 Advanced Hybrid and Ensemble Architectures

Future research should explore architectures that combine the temporal sequence modelling of LSTM/GRU with the multi-scale decomposition capability of VMD/CEEMDAN, the feature selection efficiency of LightGBM and the long-horizon dependence modelling of Transformer self-attention. Ensemble methods that dynamically weight component models based on rolling forecast error—including stacking meta-learners and dynamic mixture of experts—are particularly promising for handling the non-stationarity characteristic of Indian load series.

9.2 IoT Integration and Smart Grid Data Ecosystems

India's progressive deployment of smart meters under the RDSS scheme—targeting 250 million smart meters by 2025—creates an unprecedented opportunity to leverage high-resolution, consumer-level consumption data for ML-based forecasting. Integration of smart meter streams with IoT weather sensor networks, building management system data and industrial SCADA archives can enable bottom-up, consumer-segment-specific forecasting with substantially improved aggregate accuracy. The Scientific Reports spatial-temporal GRU study [44] and related spatio-temporal graph network architectures represent methodological pathways for multi-node smart grid forecasting.

9.3 Foundation Models and Transfer Learning

Foundation models pre-trained on large corpora of diverse time series data—such as Google's TimesFM, Amazon's Chronos and Microsoft's TiDE—offer zero-shot or few-shot forecasting capabilities that may be particularly valuable for data-scarce regional applications. Transfer learning approaches, where models pre-trained on European or North American grid data are fine-tuned on Indian utility data, could provide strong initialisation and reduce the training data volume requirements for Maharashtra-specific models. The deep learning survey by Gasparin et al. [23] identifies transfer learning as a high-priority future direction for STLF.

9.4 XAI-Enhanced Regulatory Deployment

Development of XAI-enhanced forecasting frameworks that produce SHAP-based feature attribution reports, temporal attention visualisations and natural language summaries of forecast drivers represents a critical pathway to operational deployment in Indian utilities. Such frameworks would support regulatory reporting to CERC and MERC, facilitate peer review of forecast methodologies and enable non-specialist users to interrogate model behaviour. Baur et al. [42] identify operational viability as the least-addressed dimension of XAI in energy forecasting and this gap is especially acute in the Indian regulatory context.

9.5 Adaptive Online Learning Systems

Online learning architectures that continuously update model parameters as new observations become available—without full retraining—are essential for managing concept drift in rapidly evolving demand environments. Meta-learning and continual learning approaches can enable models to rapidly adapt to new seasonal regimes, post-event behavioural shifts, or major changes in industrial load composition. The Bayesian DA-LSTM framework [45] proposed drift-adaptive interval load forecasting with LSTM networks, demonstrating significant improvements in MAPE stability over extended deployment periods.

10. Conclusion

This paper has offered a thorough critical review of machine and deep learning methods for electricity demand forecasting, with the analysis based on research published in 2018-2024 in IEEE Transactions, Applied Energy, Scientific Reports, Frontiers in Energy Research and other major Scopus-indexed journals. The conclusion is clear: ML approaches, especially hybrid deep learning models which combine signal decomposition, recurrent networks and attention mechanisms, significantly outperform conventional statistical methods, with a MAPE of less than 1.5% under the right circumstances on standard datasets.

Of the individual models, LSTM and GRU recurrent networks offer the best combination of forecasting accuracy, temporal modelling and operational readiness. CNN-LSTM architectures with attention mechanisms are the current best practice, achieving MAPE of 1.2-1.6% using a variety of international datasets. Attention-based transformer models such as the Informer and TFT demonstrate impressive results for long-term prediction and have the added advantage of interpretability. Tree-based models (XGBoost and LightGBM) continue to perform well for medium-term forecasts, while being preferred for considerations of interpretability and speed of training.
The review identified five key research gaps: the lack of state-specific ML forecasting models for Indian states such as Maharashtra; inadequate consideration of macroeconomic variables; lack of XAI approaches; almost exclusive focus on point forecasts; and lack of user-friendly forecasting platforms for the utilities. Closing these gaps calls for ongoing research effort that integrates ML methods development with domain expertise of the regulatory, weather and socioeconomic context of India's power sector.

The gradual roll-out of smart meters under the RDSS program in India, together with the country's rapidly strengthening capability in ML research and the growing regulatory support for data-driven grid operation, provides an excellent opportunity for effective research in this area. Precise, interpretable and localized electricity demand forecasting is therefore not just a technical, but a social and developmental challenge - enabling better resource allocation, decarbonisation and energy security for the Indian economy.

References

[1] M. Q. Raza and A. Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," Renewable and Sustainable Energy Reviews, vol. 50, pp. 1352–1372, 2015. doi: 10.1016/j.rser.2015.04.065

[2] A. Dedinec, S. Filiposka, A. Dedinec and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of Macedonian case," Energy, vol. 115, pp. 1688–1700, 2016. doi: 10.1016/j.energy.2016.07.090

[3] Central Electricity Authority, Annual Report 2022–23, Ministry of Power, Government of India, New Delhi, 2023. [Online]. Available: https://cea.nic.in

[4] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," International Journal of Forecasting, vol. 32, no. 3, pp. 914–938, 2016. doi: 10.1016/j.ijforecast.2015.11.011

[5] L. Hernandez et al., "A survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings," IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1460–1495, 2014. doi: 10.1109/SURV.2014.032014.00094
[6] H. S. Oliveira and H. P. Oliveira, "Transformers for energy forecast," Sensors, vol. 23, no. 15, p. 6840, Aug. 2023. doi: 10.3390/s23156840
[7] H. Shi, M. Xu and R. Li, "Deep learning for household load forecasting: A novel pooling deep RNN," IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5271–5280, Sep. 2018. doi: 10.1109/TSG.2018.2816089
[8] Z. Zhang, D. Dong and X. Liu, "Short-term load forecasting based on CEEMDAN and transformer," Electric Power Systems Research, vol. 214, p. 108885, Jan. 2023. doi: 10.1016/j.epsr.2022.108885
[9] J. Lin, J. Ma, J. Zhu and Y. Cui, "Short-term load forecasting based on LSTM networks considering attention mechanism," International Journal of Electrical Power & Energy Systems, vol. 137, p. 107818, May 2022. doi: 10.1016/j.ijepes.2021.107818
[10] J. W. Chan and C. K. Yeo, "A transformer based approach to electricity load forecasting," The Electricity Journal, vol. 37, no. 2, Mar. 2024. doi: 10.1016/j.tej.2024.107385
[11] G. E. P. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ: John Wiley & Sons, 2015.
[12] R. Kumar, P. Kumar and Y. Kumar, "Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms," International Journal of Information Technology, vol. 14, no. 1, pp. 359–373, 2022. doi: 10.1007/s41870-021-00585-8
[13] C. E. Velasquez, M. Zocatelli, F. B. Estanislau and V. F. Castro, "Analysis of time series models for Brazilian electricity demand forecasting," Energy, vol. 247, p. 123483, May 2022. doi: 10.1016/j.energy.2022.123483
[14] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil and M. N. S. Khan, "Regression model-based short-term load forecasting for university campus load," IEEE Access, vol. 10, pp. 8891–8905, 2022. doi: 10.1109/ACCESS.2022.3144206
[15] V. N. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. New York: Springer, 2000. doi: 10.1007/978-1-4757-3264-1
[16] S. C. Gupta et al., "Evaluation of electrical load demand forecasting using various machine learning algorithms," Frontiers in Energy Research, vol. 12, p. 1408119, May 2024. doi: 10.3389/fenrg.2024.1408119
[17] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324
[18] C. Sekhar and R. Dahiya, "Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand," Energy, vol. 268, p. 126660, Apr. 2023. doi: 10.1016/j.energy.2023.126660
[19] X. Deng, L. Ye, J. Li, M. Liu and P. Huang, "Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model," Energy Reports, vol. 8, pp. 8661–8674, Nov. 2022. doi: 10.1016/j.egyr.2022.06.053
[20] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785
[21] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 3146–3154.
[22] Y. Wang, J. Chen, X. Chen, X. Zeng, Y. Kong, S. Sun, Y. Guo and Y. Liu, "Short-term load forecasting for industrial customers based on TCN-LightGBM," IEEE Transactions on Power Systems, vol. 36, no. 3, pp. 1984–1997, May 2021. doi: 10.1109/TPWRS.2020.3028133
[23] A. Gasparin, S. Lukovic and C. Alippi, "Deep learning for time series forecasting: The electric load case," CAAI Transactions on Intelligence Technology, vol. 7, no. 1, pp. 1–25, Mar. 2022. doi: 10.1049/cit2.12060
[24] A. Wan, Q. Chang, A.-B. Khalil and J. He, "Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism," Energy, vol. 282, p. 128274, Nov. 2023. doi: 10.1016/j.energy.2023.128274
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. doi: 10.1162/neco.1997.9.8.1735
[26] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841–851, Jan. 2019. doi: 10.1109/TSG.2017.2753802
[27] S. Bouktif, A. Fiaz, A. Ouni and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm," Energies, vol. 11, no. 7, p. 1636, Jun. 2018. doi: 10.3390/en11071636
[28] J. Lin, J. Ma, J. Zhu and Y. Cui, "Short-term load forecasting based on LSTM networks considering attention mechanism," International Journal of Electrical Power & Energy Systems, vol. 137, p. 107818, May 2022. doi: 10.1016/j.ijepes.2021.107818
[29] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proc. EMNLP 2014, Doha, Qatar, Oct. 2014, pp. 1724–1734. doi: 10.3115/v1/D14-1179

[30] J. Chen, L. Liu, K. Guo, S. Liu and D. He, "Short-term electricity load forecasting based on improved data decomposition and hybrid deep-learning models," Applied Sciences, vol. 14, no. 13, p. 5966, 2024. doi: 10.3390/app14135966
[31] N. Ahmed and F. Amin, "Deep learning-driven hybrid model for short-term load forecasting and smart grid information management," Scientific Reports, vol. 14, p. 13548, Jun. 2024. doi: 10.1038/s41598-024-63262-x
[32] P. Ran, K. Dong, X. Liu and J. Wang, "Short-term load forecasting based on CEEMDAN and transformer," Electric Power Systems Research, vol. 214, p. 108885, Jan. 2023. doi: 10.1016/j.epsr.2022.108885
[33] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.
[34] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in Proc. AAAI Conference on Artificial Intelligence, vol. 35, no. 12, Virtual, Feb. 2021, pp. 11106–11115. doi: 10.1609/aaai.v35i12.17325
[35] Z. Wang, Z. Zhu, G. Xiao, B. Bai and Y. Zhang, "A Transformer-based multi-entity load forecasting method for integrated energy systems," Frontiers in Energy Research, vol. 10, p. 952420, Jun. 2022. doi: 10.3389/fenrg.2022.952420
[36] B. Lim, S. O. Arik, N. Loeff and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748–1764, 2021. doi: 10.1016/j.ijforecast.2021.03.012
[37] D. Niu, M. Yu, L. Sun, T. Gao and K. Wang, "Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism," Applied Energy, vol. 313, p. 118801, May 2022. doi: 10.1016/j.apenergy.2022.118801
[38] W. H. Chung, Y. H. Gu and S. J. Yoo, "District heater load forecasting based on machine learning and parallel CNN-LSTM attention," Energy, vol. 246, p. 123350, May 2022. doi: 10.1016/j.energy.2022.123350
[39] H. Shahinzadeh et al., "Review and comparative analysis of deep learning techniques for smart grid load forecasting," in Proc. 2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), IEEE, 2024, pp. 1–9. doi: 10.1109/AISP61396.2024.10475266
[40] Y. Ahmed, M. Khalid, A. H. Mazhar and R. Awan, "A hybrid model based on machine learning and time series decomposition for short-term electricity load forecasting," IEEE Access, vol. 10, pp. 12497–12510, Jan. 2022. doi: 10.1109/ACCESS.2022.3145452
[41] Y. Yang, W. Li, T. A. Gulliver and S. Li, "Bayesian deep learning-based probabilistic load forecasting in smart grids," IEEE Transactions on Industrial Informatics, vol. 16, no. 7, pp. 4703–4713, Jul. 2020. doi: 10.1109/TII.2019.2942353
[42] L. Baur, K. Ditschuneit, M. Schambach, C. Kaymakci, T. Wollmann and A. Sauer, "Explainability and interpretability in electric load forecasting using machine learning techniques – a review," Energy and AI, vol. 15, p. 100328, Mar. 2024. doi: 10.1016/j.egyai.2023.100328
[43] M. Arrieta-Prieto and C. Schell, "AI explainability and governance in smart energy systems: A review," Frontiers in Energy Research, vol. 10, p. 1071291, Jan. 2023. doi: 10.3389/fenrg.2023.1071291
[44] Z. Wang et al., "Electricity consumption prediction using an advanced spatial-temporal deep learning framework," Scientific Reports, vol. 15, p. 14892, Apr. 2025. doi: 10.1038/s41598-026-46825-y
[45] F. Bayram, P. Aupke, B. S. Ahmed, A. Kassler, A. Theocharis and J. Forsman, "DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks," Engineering Applications of Artificial Intelligence, vol. 123, p. 106460, Aug. 2023. doi: 10.1016/j.engappai.2023.106460