

A Unified Comparative Framework for Short-Term Load Forecasting: Multi-Paradigm Machine Learning Evaluation with Weather-Integrated Features

Esther Omoyiwola

*Electrical and Computer Engineering
Georgia Southern University*

Statesboro, Georgia

eo03837@georgiasouthern.edu

Roland Kobla Tagayi

*Electrical and Computer Engineering
Georgia Southern University*

Statesboro, Georgia

rolandtagayi@gmail.com

Yvonne Okafor

*Computer Science
Georgia Southern University*

Statesboro, Georgia

yo00569@georgiasouthern.edu

Akinfenwa Ayobami

*Mechanical Engineering
Georgia Southern University
Statesboro, Georgia
akinfenwaas@gmail.com*

Abstract—Short-term electricity load forecasting (STLF) is essential for reliable grid operation, economic dispatch, and energy management in modern power systems characterized by renewable integration and climate-driven variability. Over the past decades, forecasting methodologies have evolved from traditional statistical techniques to machine learning and deep learning architectures. However, inconsistencies in benchmarking practices, preprocessing pipelines, and weather feature integration limit the interpretability and comparability of reported results. This study presents a structured literature review and proposes a unified comparative evaluation framework for short-term electricity load forecasting. The review synthesizes traditional statistical models, machine learning approaches, and deep learning architectures, with particular emphasis on the integration of meteorological variables. Furthermore, the study highlights the need for systematic weather impact sensitivity analysis to quantify the marginal contribution of environmental features to predictive performance across model classes.

By identifying methodological gaps and inconsistencies in existing studies, this work establishes a reproducible foundation for fair model comparison and weather-aware forecasting design. The proposed framework aims to bridge model performance evaluation with environmental dependency analysis, contributing toward more robust and practically applicable load forecasting systems.

Index Terms—Energy load forecasting, machine learning, data science, time-series analysis

I. INTRODUCTION

ACCURATE short-term electricity load forecasting (STLF) is a foundational requirement for modern power system operation. In electricity markets and vertically integrated utilities alike, reliable load prediction enables efficient unit commitment, economic dispatch, reserve allocation, and real-time balancing. Inaccurate forecasts

can lead to over-generation, increasing operational costs, or under-generation, risking system instability and service interruptions. With increasing penetration of renewable energy sources and demand-side variability, the forecasting problem has become more complex, nonlinear, and weather-sensitive. Over the past decades, forecasting methodologies

have evolved from traditional statistical techniques such as ARIMA and SARIMA to machine learning (ML) models and, more recently, deep learning (DL) architectures including recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. While these approaches demonstrate promising accuracy improvements, the literature reveals inconsistencies in benchmarking practices, feature engineering pipelines, and evaluation metrics. Moreover, although meteorological variables are commonly incorporated into forecasting models, limited studies systematically analyze the sensitivity and marginal impact of weather features across different modeling paradigms. This literature

review aims to provide a structured and critical synthesis of forecasting methodologies relevant to short-term electricity load prediction. The scope of this review includes traditional statistical models, machine learning-based approaches, deep learning architectures, and studies integrating weather variables into forecasting frameworks. Particular emphasis is placed on identifying research gaps related to unified model comparison and structured weather impact analysis, which form the foundation of the proposed study.

II. LITERATURE REVIEW

A. Foundations and Structural Context of Short-Term Electricity Load Forecasting

Short-term electricity load forecasting (STLF) remains one of the most extensively studied and operationally critical problems in modern power system engineering. Accurate short-horizon demand prediction supports unit commitment, economic dispatch, reserve scheduling, congestion management, and real-time balancing in both regulated and deregulated electricity markets. As power systems transition toward high renewable penetration and distributed energy integration, the complexity of net load behavior has increased substantially, amplifying the technical challenges associated with reliable short-term forecasting [1].

Electricity demand exhibits multi-scale temporal dependencies characterized by strong intra-day, weekly, and seasonal periodicity, superimposed with stochastic variations driven by meteorological conditions, socio-economic activity, and behavioral consumption patterns. Unlike conventional stationary time-series processes, modern electricity load profiles are influenced by nonlinear interactions between weather variables, calendar effects, distributed energy resources, and emerging electrification trends such as electric vehicles and demand response programs. These dynamics introduce non-stationarity, heteroscedasticity, and structural breaks that complicate model generalization and stability.

Historically, statistical methods such as Autoregressive integrated moving average (ARIMA) [2], multiple linear regression, and exponential smoothing formed the methodological backbone of STLF research. While effective under assumptions of linear dependence and stable seasonal patterns, these models demonstrate limited capacity to capture nonlinear feature interactions and abrupt demand shifts in highly dynamic grid environments. The limitations of purely statistical approaches motivated the transition toward data-driven machine learning paradigms capable of modeling complex, high-dimensional relationships.

Over the past decade, ensemble learning techniques (e.g., Random Forest, Gradient Boosting), kernel-based methods (e.g., Support Vector Regression), and deep learning architectures (e.g., LSTM, CNN-LSTM hybrids) have been extensively investigated for STLF applications [3]. These models help to improve representational flexibility and nonlinear modeling capacity; however, literature surveys consistently highlight persistent challenges including model overfitting, interpretability limitations, inconsistent preprocessing pipelines, and sensitivity to forecast horizon selection.

Despite significant methodological advancements, many existing studies prioritize marginal improvements in aggregate error metrics without systematically evaluating model robustness, stability across multiple temporal horizons, and the quantified contribution of exogenous weather variables under unified experimental conditions. As renewable variability and demand-side complexity continue to intensify, there remains a pressing need for structured, reproducible, and operationally

grounded comparative analyses that extend beyond single-model performance reporting.

1) *Evolution from Statistical to Machine Learning Approaches:* The methodological evolution of short-term electricity load forecasting has progressed from classical statistical time-series modeling to advanced machine learning and deep learning paradigms. Early forecasting frameworks were dominated by linear regression, autoregressive integrated moving average (ARIMA), and exponential smoothing techniques, which assume linear dependencies and relatively stable seasonal structures. While effective under structured conditions, these models often exhibit limited adaptability to nonlinear demand behavior and complex feature interactions [4].

To overcome these limitations, machine learning models were increasingly adopted for STLF applications. Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), and many more demonstrated superior nonlinear modeling capability and improved robustness to multicollinearity. Ensemble-based techniques in particular gained prominence due to their ability to reduce variance and improve generalization performance across diverse load profiles [5].

In recent years, deep learning architectures have further transformed the STLF landscape. Recurrent neural networks (RNN), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), and hybrid CNN-LSTM architectures have been extensively applied to capture temporal dependencies and sequential patterns in electricity demand data. Comprehensive reviews highlight that deep learning models outperform traditional statistical approaches in high-resolution and large-scale datasets, particularly under conditions of strong nonlinearity and weather dependency.

However, despite demonstrated accuracy improvements, recent surveys emphasize persistent challenges, including interpretability limitations, high computational cost, hyperparameter sensitivity, and model instability across seasonal regimes. These observations suggest that model selection for STLF should not rely solely on average error minimization but instead consider robustness, scalability, and operational usability.

2) *Structural Characteristics and Complexity of Electricity Load:* Electricity demand exhibits a multi-scale temporal structure characterized by strong periodic patterns at daily, weekly, and seasonal levels. Intra-day variations are typically driven by human activity cycles, while weekly and seasonal trends reflect industrial schedules and climatic shifts. These recurring structures introduce deterministic components that can be partially modeled through trend and seasonality decomposition techniques.

However, beyond deterministic periodicity, electricity load demonstrates significant nonlinear and stochastic behavior. Meteorological variables such as temperature and humidity strongly influence demand levels, particularly in regions with high heating and cooling requirements. The relationship between temperature and electricity consumption is inherently nonlinear, often exhibiting threshold effects and asymmetric responses during extreme weather conditions.

Furthermore, modern power systems experience non-stationarity due to evolving consumption patterns, distributed energy resource integration, electric vehicle integration, and demand response programs. Structural changes in demand profiles challenge assumptions of stable statistical distributions and complicate model generalization across time horizons [6].

Peak demand events introduce additional forecasting difficulty. Extreme load conditions frequently occur under abnormal weather or system stress scenarios, where even small prediction errors can have disproportionately large operational consequences. Consequently, short-term load forecasting must ensure not only average accuracy but also robustness under high-impact conditions.

3) Implications for Forecasting Methodology: The structural characteristics of electricity demand directly influence the design and selection of forecasting methodologies. Multi-scale periodicity requires models capable of capturing temporal dependencies across multiple horizons, while nonlinear weather sensitivity demands flexible functional forms that can represent complex feature interactions. Linear modeling frameworks, although effective for stable seasonal components, often struggle to accommodate threshold effects and asymmetric demand responses observed under extreme climatic conditions.

Furthermore, the presence of non-stationarity and structural evolution in load profiles imposes additional constraints on forecasting systems. Models must maintain generalization capability under shifting demand regimes and avoid overfitting to transient patterns. The increasing integration of distributed energy resources and electrified loads further amplifies volatility, reducing the effectiveness of rigid parametric approaches.

These methodological challenges have progressively shifted research focus toward adaptive and data-driven forecasting frameworks that can capture nonlinear dependencies and temporal dynamics more effectively than traditional statistical models. However, as model flexibility increases, issues of interpretability, computational complexity, and robustness become equally important considerations in practical deployment.

Therefore, the forecasting problem in modern power systems extends beyond pure error minimization. It requires a balanced evaluation of accuracy, stability, and operational reliability under evolving system conditions.

Traditional statistical approaches have historically formed the backbone of STL. Before the widespread adoption of machine learning, utilities relied heavily on time-series models such as Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and ARIMA, including seasonal extensions such as Seasonal Autoregressive Integrated Moving Average (SARIMA). These models are grounded in statistical theory and assume that future electricity demand can be expressed as a linear function of past observations and stochastic error terms. Their interpretability, relatively low computational burden, and strong theoretical foundations made them attractive for operational grid forecasting, particularly in structured and moderately stationary environments.

A recent systematic study by [7] explicitly frames STL as a multi-level problem, from the system level down to low-voltage secondary substations, and proposes an interpretable step-by-step modeling workflow that incorporates historical load, calendar effects, and weather variables as structured predictors. Their analysis highlights that even when the final forecasting framework is not purely ARIMA/SARIMA, the classical explicit seasonal, cycle modeling, and carefully defined exogenous drivers still provide strong value for operational clarity and reproducibility.

Recent STL studies consistently show that weather introduces nonlinear effects, threshold behavior, and system shifts in cooling and heating dominance. In practice, this means performance gains depend not only on including temperature and humidity, but on representing their relationships in ways the model can learn robustly. [7] emphasizes structured variable modeling with weather and calendar factors to preserve interpretability while improving accuracy across grid levels, supporting the idea that weather integration requires careful design to avoid unstable relationships across locations and seasons.

Recent empirical comparisons show where pure ARIMA baselines struggle. For example, [8] directly compares ARIMA against an ANN-based approach for load forecasting, using ARIMA as the statistical benchmark and evaluating error primarily through Mean Absolute Percentage Error (MAPE). Their results reinforce a common finding in contemporary STL: linear statistical dependence can be competitive when patterns are stable, but accuracy can degrade when the signal is affected by nonlinear load changes and shifting conditions, which is frequently the case in real demand.

[9] proposes a weather-sensitive forecasting framework based on dynamic mode decomposition with control (DMDc)—a reduced-order, data-driven dynamic modeling technique that is still closely aligned with the statistical modeling explicit dynamics, state evolution, and interpretable structure, while allowing exogenous influence through the control component. This is necessary because it shows how research is increasingly positioning classical statistical baselines alongside modern dynamic formulations that better tolerate weather-driven variability.

On the modeling side, [9]'s weather-sensitive framework is particularly relevant because it treats weather influence through a control-style mechanism (DMDc), aligning with the broader point that weather can act like an external forcing term on load dynamics. This perspective is helpful when you justify why conventional seasonal models can fail during abnormal conditions if the underlying dynamics shift, and seasonality alone is not enough.

In parallel, other studies increasingly use deep learning models either to incorporate weather more effectively or even to jointly predict load and weather variables. [10] presents a ResNet-LSTM approach that explicitly targets short-term load forecasting together with associated weather variable prediction, reinforcing the idea that weather and demand may be modeled as coupled signals rather than completely

independent inputs. This supports why integrating meteorological features is now standard practice in competitive STLF studies—and why comparing across machine learning models is necessary.

Recent reviews focused on short-term forecasting, which reinforces that weather integration is typically beneficial, but the magnitude of improvement depends on feature engineering, data alignment, hourly matching, and lag design, and the model's ability to express nonlinearities issues that become central when you compare traditional statistical baselines to ML and LSTM models [11].

B. Machine Learning Models for Power Load Forecasting

Recent research in power load forecasting has seen a considerable shift toward machine learning (ML) algorithms, largely because they have exhibited the most reliable accuracy in predicting power demands [12]. Traditional statistical models frequently struggle to capture the intricate, nonlinear interactions between variables such as time and temperature. Machine learning techniques are preferred because of their ability to capture these nonlinear relationships; they can interpret calendar markers as external inputs, allowing them to better comprehend and predict power load demands [13]. Systematic data suggests that forecasting accuracy is heavily horizon-dependent. Shorter horizons often have a lower error, whereas medium to long-term forecasting is more uncertain and error-prone due to compounding behavioral and economic unpredictability [14].

Modern reviews underscore that forecasting performance is determined by an amalgamation of model type, feature engineering, and dataset properties. For example, a model designed for consistent grid-level demand may struggle with the fluctuating "noise" of household smart meters. Ultimately, success requires precisely matching the analytical tools and data preparation to the specific scale of the problem. Vivas, et al. [13] emphasize that the majority of published studies remain site-specific, making cross-study generalization difficult and limiting reproducibility [13]. Similarly, Azeem, et al. [15] highlight that forecasting model accuracy varies significantly among sectors (residential, commercial, industrial) and generation modalities, especially when renewables are included in the supply system [15]. This reinforces the conclusion that model superiority cannot be claimed universally without validation across diverse datasets.

C. Regression-Based and Econometric Models as Baselines

Although modern forecasting literature is dominated by newer Machine Learning algorithms, regression-based models are still extensively employed as baselines due to their interpretability and low computational cost. Recent research confirms that regression techniques are especially successful for long-term load forecasting (LTLF), where demand follows smoother seasonal and economic cycles [14]. However, regression models perform poorly in short-term load forecasting (STLF) because they cannot effectively represent nonlinear

weather-load interactions or abrupt behavioral fluctuations [16], [17].

In the findings of Zhang, et al. [18], regression models are commonly used in load forecasting situations, but they are typically outperformed when nonlinear learners are incorporated, particularly when the dataset incorporates complicated operational conditions and high-frequency sensor observations [18]. In a similar vein, Vivas, et al. [13] classify regression as a classical statistical/mathematical forecasting, highlighting that while regression is still relevant, it rarely achieves state-of-the-art performance when evaluated using MAPE across multiple geographic regions [13].

Therefore, regression-based approaches are still most effective when interpretability is crucial, and forecasting jobs require coarse temporal granularity (monthly/annual), but they are insufficient for high-resolution STLF applications.

D. Support Vector Regression (SVR) and Kernel-Based Learning

Support Vector Regression (SVR) continues to be widely adopted in load forecasting research due to its capacity to model nonlinear demand relationships through kernel transformations. Recent literature indicates that SVR is frequently used in STLF applications, particularly in cases where datasets are moderate in size and the feature space is carefully engineered (manually selected, cleaned, and transformed) [17]. In microgrid forecasting contexts, SVR is often reported as a competitive benchmark due to its stability and strong performance under limited training data [19].

However, emerging evidence suggests that SVR performance deteriorates as dataset size and volatility grow. Zhang, et al. [18] report that while SVR remains a popular choice for predicting building energy consumption, it is frequently surpassed by ensemble tree-based approaches such as Random Forest or XGBoost, particularly when the model is presented with larger, more complicated datasets [18]. Furthermore, Azeem, et al. [15] establish that the SVR predicting error differs dramatically across generating modalities. SVR, in particular, performs poorly when a large number of renewable energy sources (such as solar or wind) are added to the mix, because renewable-driven intermittency results in nonstationary demand patterns that SVR struggles to process [15].

A consistent theme is that SVR performs well in controlled STLF scenarios but lacks scalability and adaptability relative to modern ensemble-based ML frameworks.

E. Tree-Based Models: Decision Trees and Random Forests (RF)

Tree-based learning methods, particularly Random Forest (RF), continue to be among the most widely reported ML strategies for load forecasting due to their ability to capture nonlinear feature interactions without relying on strong parametric assumptions. Multiple studies evaluate RF as a high-performing and robust method for demand forecasting, particularly when climatic data and lagged load terms are included [15], [16].

The main advantage of RF over single decision trees is its resistance to overfitting, as well as its ability to quantify feature importance, which improves interpretability in operational planning. Zhang, et al. [18] report RF as one of the most often utilized ML algorithms in building load prediction tasks, owing to its ability to balance accuracy with model transparency [18]. Similarly, Vivas, et al. [13] identify RF as one of the most frequently appearing ML models among studies reporting MAPE performance results, demonstrating its broad adoption across geographic forecasting applications [13].

However, research suggests that boosting-based ensembles frequently outperform RF performance when the forecasting challenge needs fine-grained error reduction. This shows that RF is better described as a strong benchmark and practical deployment model, but not always the most accurate model in competitive forecasting situations [16], [18].

F. Boosting Models: XGBoost and Gradient Boosting Machines

Boosting-based ensemble learning, namely gradient boosting decision trees, has emerged as the dominating algorithm in modern load forecasting research. According to multiple evaluations, boosting models outperform both linear regression and RF in STLF tasks because of their capacity to iteratively correct prediction residuals and capture high-order nonlinear relationships [15], [16]. Habbak, et al. [16] identified boosting algorithms as among the most effective AI-based forecasting families in smart grid situations, owing to their reliance on structured feature representations such as temperature, time-of-day, and weekday/holiday signals [16]. Similarly, Zhang, et al. [18] describe boosting-based approaches as highly effective in developing load prediction models, especially when datasets include high-dimensional metadata and engineered time-series lags [18]. Vivas, et al. [13] bolster this conclusion by reporting that ML models, including exogenous variables, regularly yield lower MAPE scores than models that depend solely on historical load values [13].

However, boosting methods introduce challenges in hyperparameter sensitivity and reproducibility. Many studies optimize boosting architecture on single datasets, limiting the external validity of their results. Moreover, model interpretability remains weaker than regression baselines, despite the availability of feature importance tools.

Overall, boosting models represent one of the strongest power load forecast approaches, particularly for utility-scale STLF with structured exogenous predictors.

G. Hybrid ML Systems and Optimization-Assisted Forecasting Pipelines

An emerging trend in forecasting research is the increasing use of hybrid ML pipelines that integrate feature selection, optimization algorithms, and decomposition strategies. Rather than relying solely on the predictive learner, studies frequently incorporate preprocessing stages such as clustering, noise

reduction, or heuristic parameter tuning to enhance forecasting stability [13], [15].

Azeem, et al. [15] emphasize that modern forecasting systems increasingly combine clustering, optimization, and machine learning to address sector-specific forecasting demands, particularly in industrial and isolated-grid forecasting settings where demand patterns vary significantly from conventional residential loads [15]. In a similar vein, Zhang, et al. [18] highlight that feature engineering and preprocessing are now treated as core determinants of forecasting performance rather than optional steps [18]. This aligns with the systematic observations by Vivas, et al. [13], who argue that performance improvements in recent literature are often driven by the inclusion of exogenous variability sources rather than the learning model alone [13].

Despite strong reported performance, hybrid models frequently suffer from limited transparency and weak generalization. Many studies fail to assess if hybrid advantages hold up with distribution variations such as new seasons, policy changes, or renewable penetration. This suggests that hybrid forecasting systems, while powerful, are under-validated in deployment-representative situations.

H. Model Performance Reporting and Validation Limitations

Modern load forecasting studies predominantly report performance using MAPE, RMSE, and MAE, with MAPE being particularly common due to its comparability across datasets of different scales [13]. However, publication quality issues remain substantial. Vivas, et al. [13] explicitly emphasize that forecasting studies are frequently difficult to compare due to differences in time horizon, regional conditions, and dataset resolution [13]. Likewise, Zhang, et al. [18] highlight that the lack of standardized benchmarking datasets and inconsistent reporting of preprocessing steps reduces reproducibility and weakens cross-study synthesis [18].

Another critical gap is the limited use of external validation. Many studies validate performance only through internal dataset splits, which can overestimate real-world deployment performance. This is particularly problematic for smart grid forecasting, where demand behavior shifts significantly under extreme weather events, renewable intermittency, and evolving economic conditions [15], [16].

Therefore, while modern ML forecasting accuracy appears improved relative to earlier approaches, the reliability of reported gains is often constrained by limited generalization testing.

I. Comparative Synthesis of ML Model Suitability

Synthesizing evidence suggests that model suitability follows a clear hierarchy based on forecasting horizon and dataset characteristics. The comparative synthesis of models along with their strengths and limitations is summarized in Table I.

III. METHODOLOGY OF LITERATURE REVIEW

To ensure a rigorous and systematic review process, a structured methodology was adopted for identifying, selecting, and analyzing relevant scholarly articles.

TABLE I
COMPARATIVE SYNTHESIS OF ML MODEL SUITABILITY IN POWER LOAD FORECASTING

Model Category	Suitable Forecasting Horizon	Dataset Characteristics	Strengths	Limitations	Key Refs
Regression-Based Models	Long-Term Load Forecasting (LTLF)	Aggregated demand; smoother seasonal/economic trends; limited nonlinear volatility	High interpretability; low computational cost; stable for coarse temporal granularity	Poor handling of nonlinear weather-load interactions; weak STL forecast performance	[18], [19]
Support Vector Regression (SVR)	Short-Term Load Forecasting (STLF)	Moderate-sized datasets; engineered nonlinear features	Strong nonlinear modeling; stable under limited training data	Scalability issues; performance declines in large or volatile datasets; often outperformed by ensembles	[15], [18]
Random Forest (RF)	STLF and Medium-Term LF	Medium-scale datasets; structured meteorological and lagged features	Robust to overfitting; interpretable via feature importance; consistent performance	May be surpassed by boosting models in fine-grained error optimization	[13], [18]
Boosting Models (e.g., Gradient Boosting)	Primarily STLF	Large datasets; structured exogenous predictors (weather, calendar effects)	High predictive accuracy; strong nonlinear and high-order interaction modeling	Hyperparameter sensitivity; reduced interpretability; reproducibility challenges	[16], [19]
Hybrid ML Pipelines	STLF (complex/volatile environments)	High-resolution, volatile, or renewable-integrated datasets	Improved accuracy through preprocessing, decomposition, and optimization integration	Dataset-specific gains; weak external validation; reduced transparency	[13], [15]

A. Databases Accessed

The primary databases used for literature retrieval included:

- IEEE Xplore Digital Library
- ScienceDirect
- SpringerLink
- Google Scholar

These databases were selected due to their extensive coverage of peer-reviewed journals and conference proceedings in power systems, machine learning, and energy analytics.

B. Keywords Used

The literature search was conducted using combinations of the following keywords:

- short-term load forecasting
- electricity demand prediction
- machine learning load forecasting
- deep learning for power systems
- weather impact on electricity load
- LSTM load forecasting
- ARIMA electricity forecasting
- hybrid forecasting models
- electricity load sensitivity analysis

Boolean operators such as AND and OR were used to refine search results.

C. Inclusion Criteria

Articles were included in the review if they satisfied the following criteria:

- Peer-reviewed journal articles or conference proceedings.
- Published between 2015 and 2024 (with selected foundational works prior to 2015).
- Focused explicitly on short-term electricity load forecasting.

- Employed statistical, machine learning, or deep learning forecasting techniques.
- Reported quantitative performance metrics (e.g., MAE, RMSE, MAPE).
- Incorporated or discussed weather-related variables in forecasting models.

D. Exclusion Criteria

Articles were excluded if they:

- Focused exclusively on long-term forecasting.
- Addressed non-electricity domains without a transferable methodology.
- Lacked methodological transparency or quantitative evaluation.
- Were non-peer-reviewed sources, such as blogs or opinion articles.

E. Article Screening Process

The initial search yielded a broad set of publications. Titles and abstracts were screened for relevance, followed by full-text analysis of shortlisted articles. Selected papers were categorized based on methodological approach (statistical, machine learning, deep learning, hybrid) and the extent of weather feature integration. Comparative attributes such as datasets used, preprocessing methods, evaluation metrics, and reported limitations were systematically extracted to support structured analysis.

IV. SYNTHESIS AND CONCLUSION

Despite the extensive body of work on short-term electricity load forecasting, several critical gaps remain in the literature. First, while numerous studies propose statistical, machine learning, or deep learning models, most comparative analyses are limited in scope. Many works evaluate only two or three

models, often under different datasets, preprocessing pipelines, or evaluation metrics, making cross-study comparisons inconclusive. The absence of a unified benchmarking framework restricts the ability to objectively assess the relative strengths and weaknesses of forecasting paradigms.

Second, although weather variables such as temperature, humidity, and wind speed are frequently incorporated into forecasting models, limited research systematically evaluates their marginal and conditional impact across different modeling paradigms. Existing studies often include weather features without explicitly quantifying their sensitivity effects or examining how different models respond to meteorological variability. Consequently, it remains unclear whether performance improvements stem from model architecture or from feature augmentation.

Third, most forecasting studies emphasize average performance metrics such as MAE or RMSE without evaluating model robustness under extreme weather conditions or peak demand events. Given that operational risk in power systems is often associated with abnormal climatic scenarios, the lack of structured weather sensitivity analysis represents a significant limitation in practical applicability.

Furthermore, inconsistencies in feature engineering practices, including lag selection, normalization strategies, and temporal encoding, introduce additional variability in reported results. Without controlled preprocessing and identical feature pipelines, model comparisons risk reflecting experimental bias rather than intrinsic predictive capability.

To address these limitations, this study proposes a systematic comparative evaluation of multiple forecasting paradigms under a unified experimental framework. Specifically, we incorporate structured weather impact sensitivity analysis to quantify the contribution of environmental variables to predictive performance across different model classes. By maintaining consistent datasets, preprocessing procedures, and evaluation metrics, this work aims to provide a fair, reproducible, and practically meaningful assessment of forecasting performance in weather-influenced electricity demand environments.

This approach contributes to the literature by bridging model comparison with feature sensitivity analysis, thereby offering deeper insight into both predictive accuracy and environmental dependency behavior in short-term load forecasting systems.

REFERENCES

- [1] K. Ullah, M. Ahsan, S. M. Hasanat, M. Haris, H. Yousaf, S. F. Raza, R. Tandon, S. Abid, and Z. Ullah, "Short-term load forecasting: A comprehensive review and simulation study with cnn-lstm hybrids approach," *IEEE Access*, vol. 12, pp. 111 858–111 881, 2024.
- [2] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 1394–1401.
- [3] Y. Xiong, W. Wang, L. Luo, H. Chen, J. Gong, and X. Wu, "Study of short-term load forecasting based on fcm and multiple woa-lssvm combination under electricity market environment," in *2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, 2022, pp. 201–205.
- [4] I. Bhattacharjee and P. Bhattacharja, "Stock price prediction: A comparative study between traditional statistical approach and machine learning approach," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, 2019, pp. 1–6.
- [5] V. Negri, A. Mingotti, R. Tinarelli, and L. Peretto, "Comparison between the machine learning and the statistical approach to the forecasting of voltage, current, and frequency," in *2023 IEEE 13th International Workshop on Applied Measurements for Power Systems (AMPS)*, 2023, pp. 01–06.
- [6] K. Mustafa, S. Khan, S. Aslam, H. Herodotou, N. Ashraf, A. Daraz, and T. Alkhalfah, "Toward holistic energy management by electricity load and price forecasting: A comprehensive survey," *IEEE Access*, vol. 11, pp. 132 604–132 626, 2023.
- [7] M. G. Pinheiro, S. C. Madeira, and A. P. Francisco, "Short-term electricity load forecasting—a systematic approach from system level to secondary substations," *Applied Energy*, vol. 332, p. 120493, Feb. 2023.
- [8] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on arima and ann approaches," *Energy Reports*, Nov. 2023.
- [9] A. Mansouri, A. H. Abolmasoumi, and A. A. Ghadimi, "Weather sensitive short term load forecasting using dynamic mode decomposition with control," *Electric Power Systems Research*, vol. 221, p. 109387, Aug. 2023.
- [10] X. Chen, W. Chen, V. A. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using resnet-lstm based deep learning," *IEEE Access*, Jan. 2023.
- [11] Y. Eren and I. Küçükdemir, "A comprehensive review on deep learning approaches for short-term load forecasting," *Renewable and Sustainable Energy Reviews*, vol. 189, p. 114031, Jan. 2024.
- [12] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, p. 103406, 2022.
- [13] E. Vivas, H. Allende-Cid, and R. Salas, "A systematic review of statistical and machine learning methods for electrical power forecasting with reported MAPE score," *Entropy*, vol. 22, no. 12, p. 1412, 2020.
- [14] M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and models for electric load forecasting: a comprehensive review," *Logistics, Supply Chain, Sustainability and Global Challenges*, vol. 11, no. 1, pp. 51–76, 2020.
- [15] A. Azeem, I. Ismail, S. M. Jameel, and V. R. Harindran, "Electrical load forecasting models for different generation modalities: a review," *IEEE Access*, vol. 9, pp. 142 239–142 263, 2021.
- [16] H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, and M. I. Ibrahim, "Load forecasting techniques and their applications in smart grids," *Energies*, vol. 16, no. 3, p. 1480, 2023.
- [17] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine learning, deep learning and statistical analysis for forecasting building energy consumption—a systematic review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105287, 2022.
- [18] L. Zhang *et al.*, "A review of machine learning in building load prediction," *Applied Energy*, vol. 285, p. 116452, 2021.
- [19] R. Wazirali, E. Yaghoubi, M. S. S. Abujazar, R. Ahmad, and A. H. Vakili, "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques," *Electric Power Systems Research*, vol. 225, p. 109792, 2023.