A Deep Learning and Metaheuristic Optimization Framework for Short-Term Electricity Consumption Forecasting Using High-Resolution SCADA Data

Wei Hong Lim[1,*], Amel Ali Alhussan[2]

[1]Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia

[2]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

Emails: limwh@ucsiuniversity.edu.my; aaalhussan@pnu.edu.sa

Abstract

Accurate prediction of electricity consumption is a critical requirement for improving operational efficiency, enhancing grid reliability, and supporting sustainability objectives in urban power distribution systems, particularly in regions experiencing steady population growth and increasing demand pressure. Motivated by the limitations of conventional statistical and physics-inspired forecasting approaches, as well as the strong sensitivity of deep learning architectures to hyperparameter configuration, this study proposes a robust data-driven framework that integrates deep learning with advanced metaheuristic optimization for high-precision short-term electricity consumption forecasting. The main contribution of this work lies in the systematic development and evaluation of hybrid metaheuristic–Bidirectional Long Short-Term Memory (BiLSTM) models, in which multiple state-of-the-art optimization algorithms are employed to tune model hyperparameters. Particular emphasis is placed on the integration of the Ninja Optimization Algorithm with BiLSTM (NijOA + BiLSTM), which is designed to effectively navigate complex, high-dimensional hyperparameter search spaces encountered in deep learning–based load forecasting tasks. Baseline experiments demonstrate that BiLSTM outperforms other deep learning models, including Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU), achieving a baseline Root Mean Squared Error (RMSE) of 0.0964 and a coefficient of determination (R^2) of 0.854. These results confirm the advantage of bidirectional temporal learning in capturing the nonlinear and time-dependent characteristics of electricity consumption recorded at high temporal resolution from SCADA systems. Following metaheuristic optimization, the NijOA + BiLSTM model delivers a substantial improvement in predictive performance. The optimized configuration reduces RMSE to 0.0038, Mean Squared Error (MSE) to 1.45×10^{-5}, and Mean Absolute Error (MAE) to 0.00019, while increasing the correlation strength to $r = 0.973$ and the explanatory power to $R^2 = 0.97$. Comparative analysis across different optimization strategies further confirms the superiority of the NijOA + BiLSTM hybrid model over alternative configurations, including WAO + BiLSTM, BBO + BiLSTM, GA + BiLSTM, SFS + BiLSTM, DE + BiLSTM, and JAYA + BiLSTM. The implications of these findings are significant for real-world urban electricity distribution applications. The proposed framework enables highly accurate and reliable short-term electricity consumption forecasting, making it well suited for deployment within smart grid and distribution management systems. Such predictive capability can support informed operational decision-making, improve demand-side management strategies, reduce uncertainty in short-term planning, and contribute to the long-term sustainability and resilience of urban power distribution networks.

Keywords: Short-term electricity consumption forecasting; Deep learning; BiLSTM; Metaheuristic optimization; SCADA data

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

1 Introduction

The global energy sector is undergoing a rapid transformation driven by increasing urbanization, population growth, climate change concerns, and rising pressure on existing power system infrastructures [1], [2]. In many developing and emerging economies, these challenges are further intensified by structural energy dependencies, limited domestic energy resources, and steadily increasing electricity demand associated with socio-economic development [3], [4]. Within this broader context, Morocco represents a particularly relevant case, as the country continues to face significant challenges related to energy security, supply reliability, and the efficient management of electricity consumption [5], [6], [7]. Morocco's national energy system is characterized by a strong reliance on imported fossil fuels, especially following the shutdown of the country's sole oil refinery in 2015 [8]. Despite continuous national efforts to diversify the energy mix and expand renewable energy capacity, the country's per-capita energy consumption remains relatively low compared to regional averages. This structural dependency places additional strain on electricity distribution systems, which must simultaneously accommodate increasing demand, ensure operational reliability, and maintain acceptable service quality. Consequently, enhancing the efficiency of electricity consumption management has become a strategic priority at both national and urban levels [9]. Urban areas play a critical role in this transition, as they concentrate population growth, economic activities, and electricity demand. Tetouan, a city located in northern Morocco along the Mediterranean coast, exhibits distinctive geographic, climatic, and demographic characteristics that directly influence electricity consumption behavior. The city experiences mild and rainy winters as well as hot and dry summers, leading to pronounced seasonal variability in electricity demand. In parallel, sustained population growth has increased pressure on existing distribution infrastructures, reinforcing the need for advanced forecasting and planning tools capable of supporting efficient network operation.

Electricity distribution in Tetouan is managed by Amendis, a public service operator responsible for the supply of low- and medium-voltage consumers since 2002. The distribution network is monitored using a Supervisory Control and Data Acquisition (SCADA) system, which provides high-resolution measurements of electricity consumption across multiple distribution zones, namely Quads, Smir, and Boussafou. Such a data-rich operational environment enables detailed analysis of consumption patterns and creates opportunities for deploying advanced data-driven forecasting methodologies [10], [11]. Accurate short-term load forecasting within this framework is essential for ensuring grid stability, optimizing dispatch decisions, and minimizing technical and operational losses. Short-term electricity load forecasting at high temporal resolution, such as ten-minute intervals, has become increasingly important for modern power system operation [12]. High-frequency forecasts support real-time decision-making, demand response strategies, preventive maintenance, and efficient resource allocation. For electricity distribution network operators, precise short-term forecasts contribute to preventing overload conditions, improving voltage regulation, and enhancing overall energy efficiency. In urban systems facing growing demand and limited expansion capacity, these capabilities are particularly critical [13], [14].

Traditional statistical forecasting approaches, although historically valuable, often struggle to capture the nonlinear, non-stationary, and multivariate nature of modern electricity consumption data. Such methods typically rely on simplifying assumptions that limit their ability to model complex temporal dependencies, especially when high-frequency measurements and multiple exogenous variables are involved. In contrast, machine learning and deep learning techniques have demonstrated strong potential for overcoming these limitations by learning complex patterns directly from data. Deep learning architectures designed for sequential data analysis, including recurrent and memory-based neural networks, have emerged as powerful tools for time-series forecasting. Models such as RNN, LSTM, GRU, and BiLSTM are particularly well suited for electricity consumption forecasting due to their ability to capture long-term temporal dependencies and nonlinear relationships between historical load values and influencing factors. These characteristics make deep learning approaches highly attractive for high-resolution electricity demand forecasting in complex urban energy systems.

Despite the advantages offered by deep learning techniques, accurate short-term electricity consumption forecasting remains a challenging task, particularly in urban environments characterized by complex and dynamic demand patterns. One of the primary challenges arises from the high-frequency nature of the available data. In this study, electricity consumption is recorded at ten-minute intervals, resulting in a large-scale time-series dataset comprising more than fifty thousand observations. While such temporal granularity provides valuable information, it also increases computational complexity and places substantial demands on

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

model training and optimization processes. Another significant challenge is associated with feature redundancy and correlation among input variables. Meteorological parameters such as temperature, humidity, and wind speed are often interrelated, and their effects on electricity consumption may overlap or interact in nonlinear ways. In addition, environmental variables such as diffuse flows introduce further complexity into the feature space. Without appropriate feature selection or dimensionality reduction mechanisms, redundant or weakly informative variables can degrade learning efficiency, increase computational cost, and negatively affect model generalization.

Hyperparameter sensitivity constitutes an additional obstacle in deep learning-based forecasting models. Architectures based on memory units, particularly LSTM-derived models, are highly dependent on the selection of hyperparameters such as learning rates, number of hidden units, batch sizes, and optimization settings. Inadequate hyperparameter configurations may result in slow convergence, unstable training behavior, or insufficient representation of temporal dependencies. Therefore, systematic and automated hyperparameter optimization strategies are essential to fully exploit the predictive capabilities of deep learning models. Overfitting and generalization issues further complicate the forecasting task. Electricity consumption patterns are influenced by seasonal variability, behavioral changes, and external factors that may not be fully represented in historical data. Developing forecasting models that maintain robust predictive performance when exposed to unseen conditions remains a central challenge, necessitating careful model design, validation strategies, and optimization mechanisms.

In light of the aforementioned challenges, the primary objective of this study is to develop a robust and scalable framework for short-term electricity consumption forecasting within an urban distribution network context. The study aims to systematically evaluate multiple neural-network-based models, including ANN, RNN, LSTM, GRU, and BiLSTM architectures, in order to assess their suitability for high-resolution electricity load forecasting using real-world SCADA data. A further objective is to investigate the integration of metaheuristic optimization techniques within the forecasting framework. These techniques are employed to enhance feature selection and hyperparameter tuning processes, thereby addressing issues related to redundancy, sensitivity, and convergence. By embedding optimization strategies into the learning pipeline, the study seeks to improve model robustness and efficiency while reducing reliance on manual parameter tuning.

Overall, the study aims to establish a methodological foundation capable of supporting accurate, stable, and computationally efficient short-term electricity consumption forecasting. The proposed framework is intended to be applicable not only to the Tetouan distribution network but also to other urban energy systems exhibiting similar operational characteristics. This study contributes to the existing literature on electricity load forecasting by proposing a hybrid framework that integrates deep learning architectures with metaheuristic optimization strategies. The primary contribution lies in the development of a metaheuristic-optimized BiLSTM-based forecasting approach tailored to high-frequency electricity consumption data collected from a real-world urban distribution network.

In addition, the study provides a structured benchmarking framework that enables systematic comparison among several deep learning models under consistent data preprocessing and evaluation conditions. This comparative perspective offers valuable insights into the relative strengths and limitations of different modeling approaches in the context of short-term load forecasting. From an applied standpoint, the proposed methodology offers practical value for smart grid operation and urban energy management, particularly in developing regions where efficient utilization of existing infrastructure is essential. The framework supports data-driven decision-making and contributes to broader efforts aimed at improving energy efficiency, reliability, and sustainability.

The remainder of this paper is organized as follows. Section 2 describes the materials and methods, including the dataset, preprocessing procedures, modeling approaches, and optimization strategies. Section 3 outlines the experimental design and evaluation methodology. Section 4 presents the analytical framework adopted for model assessment. Finally, Section 5 concludes the paper and highlights potential directions for future research.

2 Literature Review

Over the last decade, research on chiller plants has increasingly converged on a central premise: because chillers frequently constitute the single largest electricity end-use within large commercial buildings, even

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

incremental improvements in prediction accuracy and control quality can translate into meaningful reductions in operational cost and carbon emissions. Consequently, chiller-related studies have evolved from isolated performance assessment toward data-driven intelligence that supports energy-efficient operation, optimal control, and long-term system reliability. The reviewed body of work reflects this evolution by tracing a trajectory from *data-driven prediction and forecasting*, through *prediction-informed optimization and control*, and finally toward *reliability- and fault-oriented analytics* that safeguard both energy performance and equipment health.

Across this literature, two persistent engineering constraints repeatedly emerge. First, operational datasets collected from real chiller plants are often noisy, incomplete, and unevenly distributed across operating regimes, due to manual setpoints, limited load diversity, or sensor unavailability. Second, practical deployment within building automation systems requires models that remain stable under non-stationary conditions, exhibit acceptable generalization beyond observed data ranges, and operate under limited sensing while maintaining a degree of interpretability. Within these constraints, researchers have proposed a wide spectrum of machine learning and deep learning (DL) architectures, frequently reinforced through hybridization with physical knowledge, metaheuristic optimization, or domain expertise.

A foundational strand of the literature evaluates classical machine learning and mainstream deep learning architectures for predicting chiller power consumption and performing short-horizon forecasting. In a representative study conducted on an academic building in Taiwan, chiller power prediction—defined as estimating instantaneous power based on contemporaneous inputs—is benchmarked between a thermodynamic model and a multilayer perceptron (MLP), while short-term forecasting is evaluated using MLP, one-dimensional convolutional neural networks (1D-CNN), and Long Short-Term Memory (LSTM) networks [15]. The results demonstrate that, despite the availability of physics-based formulations, a sufficiently trained MLP can approximate nonlinear input–output relationships with high fidelity, highlighting the expressive power of data-driven models when operational data are adequately informative. More importantly, LSTM achieves superior performance in minute-ahead forecasting, underscoring the importance of explicitly modeling temporal dependencies in systems characterized by thermal inertia, transport delays, and control deadbands. The daily retraining strategy adopted in this work further emphasizes that chiller plants should be treated as inherently non-stationary systems, where continual model adaptation is essential for sustained accuracy rather than a mere performance enhancement [15].

These findings are reinforced by comparative studies involving Feedforward Neural Networks (FNN), Random Forests (RF), and Recurrent Neural Networks (RNN), which consistently report improved predictive accuracy when temporal structures are explicitly incorporated [16]. Such studies suggest that temporal models are not simply higher-capacity alternatives to static predictors; rather, their structural alignment with sequence dynamics allows them to better accommodate fluctuating loads, transient operating modes, and regime shifts commonly observed in real chiller operation.

Beyond baseline model comparisons, a significant portion of the literature explores hybrid modeling strategies designed to address a central limitation of purely black-box predictors: reduced robustness under extrapolation and degraded performance across wide operating envelopes. One prominent example is the hybrid Thermo-Regulated Regression Model (TRRM) combined with deep learning for coefficient of performance (COP) estimation [17]. In this framework, a physically interpretable regression model captures first-order temperature effects, while a deep neural network learns the residual error. This residual learning paradigm is conceptually important, as it partitions the modeling task into a low-dimensional, physically meaningful backbone and a data-driven refinement layer that absorbs higher-order nonlinearities and unmodeled interactions. By relieving the network from relearning obvious thermodynamic trends, the hybrid structure improves generalization and stability. The operational choice to rely exclusively on temperature measurements further reflects practical deployment realities, where flow and power sensors may be unavailable or unreliable [17].

A complementary hybridization strategy is represented by Physics-Informed Neural Networks (PINNs), which embed physical constraints directly into the learning objective or network architecture. PINN-based approaches argue that incorporating conservation laws or physical relationships can regularize training when operational data are sparse, clustered, or biased toward limited regimes, thereby improving prediction accuracy for chillers and associated subsystems such as pumps [18]. Collectively, residual learning and physics-informed learning represent two complementary responses to the same challenge: achieving robust prediction performance under data limitations typical of field-deployed chiller plants.

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Parallel to these structural hybridizations, an extensive line of research focuses on coupling deep learning models with metaheuristic optimization algorithms to enhance convergence behavior, tune hyperparameters, and improve predictive accuracy. The FFNN combined with Teaching–Learning-Based Optimization (TLBO) study highlights that chiller energy prediction performance depends not only on the choice of network architecture, but also critically on the optimizer's ability to navigate complex, non-convex parameter spaces [19]. By benchmarking TLBO against a wide range of metaheuristics—including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Differential Evolution (DE), and Ant Colony Optimization (ACO)—the study positions metaheuristic-driven tuning as a viable alternative when gradient-based or manual optimization proves unstable or computationally expensive.

This narrative is extended by studies employing hybrid CNN–LSTM architectures optimized using metaheuristics such as the Barnacles Mating Optimizer (BMO) [20]. In these models, CNN components act as temporal feature extractors and noise filters, while LSTM layers capture long-term dependencies and regime transitions. The emphasis on convergence speed, statistical validation, and reproducibility reflects practical deployment considerations, where frequent retraining and reliable performance are essential for integration into building automation workflows. The inclusion of SHAP-based explainability further signals a shift toward actionable intelligence, where model outputs must align with engineering intuition to support trust and informed decision-making [20].

A related but distinct optimization-oriented framework integrates hybrid metaheuristics with Generative Adversarial Networks (GANs) and wrapper-based feature selection [21]. Although differing in architectural details, this work shares a common assumption with other metaheuristic-based studies: raw sensor streams often contain redundancy and multicollinearity, and automated feature selection can significantly enhance generalization while reducing sensitivity to noise. When viewed alongside physics-informed and residual-learning approaches, the literature reveals two complementary philosophies for improving robustness: embedding engineering structure to constrain learning, or employing algorithmic search to discover performant configurations in high-dimensional design spaces.

Another major research direction integrates prediction models directly into control frameworks. LSTM-based predictive reinforcement learning (RL) control argues that conventional interval-based optimization often assumes quasi-static loads within control horizons—an assumption frequently violated in practice [22]. By forecasting future cooling demand and incorporating predictions into RL-based control policies, these approaches demonstrate improved energy savings compared to both rule-based control and non-predictive RL. This highlights a fundamental insight: in chiller plants, control quality is inherently dependent on forecast quality, as control decisions are evaluated over future trajectories rather than instantaneous states [22]. Complementary clustering-based RL approaches further address scalability and convergence challenges by partitioning the state space into regime-consistent subsets, thereby improving learning efficiency and stability under complex, coupled system dynamics [23]. Similar ideas are reflected in net-zero oriented studies that integrate deep learning prediction with correlation-based feature prioritization to support supervisory control design [24].

While forecasting and control address operational efficiency, reliability-focused analytics aim to preserve that efficiency under faults and degradation. Fault detection, diagnosis, and prediction have therefore emerged as critical components of intelligent chiller management. Knowledge-embedded Deep Belief Network (DBN) frameworks address multi-fault diagnosis under incomplete labeling by integrating domain rules into feature learning and classifier training [25]. Other studies focus on specific fault types, such as refrigerant undercharge, using DBN-enhanced Extreme Learning Machines (ELM) optimized by PSO and informed by expert-driven feature selection [26]. Addressing distribution shifts across working conditions, domain knowledge-assisted deep ELM frameworks explicitly target cross-condition generalization, emphasizing robustness as a primary design requirement rather than a secondary performance metric [27].

More recently, attention has shifted toward fault prediction rather than post hoc diagnosis. Modular frameworks combining autoencoders, classifiers, and adaptive thresholding illustrate early but practical steps toward predictive maintenance in HVAC systems [28]. Transfer learning studies further extend this paradigm by demonstrating that knowledge learned from source chillers can be transferred to target systems with limited data, reducing commissioning effort and accelerating deployment of analytics in real buildings.

Absorption chillers appear in the reviewed literature as a parallel domain where surrogate modeling enables accessibility and parametric exploration. Neural-network-based surrogates replacing detailed TRNSYS

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

components illustrate how deep learning can support rapid sensitivity analysis and design exploration without repeated reliance on computationally expensive simulations [29]. When considered alongside studies motivating alternative cooling technologies, this line of work underscores that decarbonization pathways involve both operational optimization of conventional systems and performance enhancement of emerging cooling technologies [21], [29].

In synthesis, the reviewed literature demonstrates a clear evolution of chiller analytics from isolated prediction tasks toward integrated frameworks that combine forecasting, optimization, control, and reliability intelligence. Temporal deep learning architectures—particularly LSTM and hybrid CNN–LSTM models—consistently emerge as effective tools for short-horizon energy forecasting under dynamic operating conditions [15], [20], [22]. However, the systematic exploration of *bidirectional* temporal architectures, such as Bidirectional LSTM (BiLSTM), remains limited in chiller energy applications, despite their potential to exploit richer temporal context and capture complex interdependencies inherent in real-world operation. At the same time, the literature clearly indicates that predictive performance and robustness are highly sensitive to hyperparameter configuration and optimizer choice, motivating the growing use of metaheuristic optimization strategies.

Despite these advances, a notable research gap persists in the systematic evaluation of advanced metaheuristic optimizers for tuning bidirectional deep temporal models in chiller energy prediction tasks under realistic field constraints, including noisy measurements, limited sensing, and non-stationary operation. Addressing this gap is essential for developing prediction frameworks that are not only accurate in retrospective evaluation, but also stable, generalizable, and suitable for deployment within practical building automation systems.

3 Materials and Methods

3.1 Dataset Description

The present study is based on a real-world electricity consumption dataset collected from the urban electricity distribution network of Tetouan city, located in northern Morocco. The data were obtained through the Supervisory Control and Data Acquisition (SCADA) system operated by Amendis, the public service concessionaire responsible for the distribution of electricity and drinking water in the Tetouan region since 2002. Amendis manages the delivery of electrical energy from the national transmission grid to end users, ensuring reliable operation of the low- and medium-voltage distribution network. The use of SCADA-based measurements guarantees high data fidelity, operational relevance, and continuous monitoring of electricity consumption under actual grid conditions, making the dataset particularly suitable for short-term load forecasting studies.

The electricity distributed within the Tetouan network is supplied by the National Office of Electricity and Drinking Water and is transformed from high voltage levels (63 kV) to medium voltage levels (20 kV) before being transported and distributed to consumers. The dataset reflects consumption at the distribution level and captures the aggregated demand behavior of residential, commercial, and service-sector consumers connected to the low- and medium-voltage network. As such, it provides a comprehensive representation of urban electricity demand dynamics rather than isolated end-user behavior.

The dataset spans a complete calendar year, from January 1 to December 31, 2017, and is recorded at a temporal resolution of ten minutes. This high-frequency sampling results in a total of 52,416 observations, enabling the detailed characterization of short-term fluctuations in electricity consumption. The chosen temporal resolution is particularly relevant for operational forecasting, as it captures rapid demand variations driven by human activity cycles, meteorological changes, and network operating conditions. Moreover, the full-year coverage ensures that seasonal effects, including winter heating demand and summer cooling loads, are fully represented in the data.

Each observation within the dataset is described by a set of input features encompassing meteorological, environmental, and electrical variables. Meteorological features include ambient temperature, relative humidity, and wind speed, which are widely recognized as influential factors in electricity consumption due to

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

their direct impact on heating, cooling, and ventilation usage. These variables provide essential contextual information for modeling weather-sensitive demand patterns in an urban environment characterized by a Mediterranean climate.

In addition to standard meteorological parameters, the dataset incorporates environmental flow variables, namely general diffuse flows and diffuse flows. These variables capture atmospheric and environmental conditions that may indirectly influence electricity consumption behavior and grid operating conditions. The inclusion of such features enriches the input space and allows the forecasting models to account for a broader range of exogenous influences beyond conventional weather indicators.

Electricity consumption data are provided separately for three major distribution zones within Tetouan city, corresponding to the Quads, Smir, and Boussafou substations. These zones represent the primary supply areas of the urban distribution network and exhibit heterogeneous consumption characteristics due to differences in population density, land use, and local demand profiles. Zone-level measurements enable the analysis of spatial variability in electricity consumption and support both localized forecasting and aggregated demand modeling. Depending on the forecasting configuration, the consumption of each zone can be modeled independently or combined to form an aggregated load signal representative of city-wide demand.

The forecasting task addressed in this study focuses on short-term electricity consumption prediction at a ten-minute resolution. Such short-term forecasts are essential for real-time grid operation, load balancing, and proactive network management. By predicting future consumption based on historical patterns and exogenous variables, the forecasting models aim to support operational decision-making within the distribution network, including congestion mitigation and efficient resource allocation.

To ensure rigorous model development and unbiased performance evaluation, the dataset was partitioned following a time-series–consistent splitting strategy. The chronological order of observations was strictly preserved to reflect realistic forecasting conditions and to prevent information leakage from future data into the training process. The dataset was divided into three mutually exclusive subsets: a training set used for model learning, a validation set employed for hyperparameter tuning and model selection, and an independent testing set reserved for final performance assessment. This partitioning approach ensures that model evaluation accurately reflects generalization capability when applied to unseen data and aligns with best practices in time-series forecasting research.

Understanding the statistical relationships among input variables is a crucial preliminary step in data-driven modeling, as it provides insight into feature dependencies, potential multicollinearity, and the relative influence of environmental and operational factors on the target system. Prior to any preprocessing or feature engineering, an exploratory correlation analysis was conducted to examine the linear associations between zone-level variables and meteorological parameters. Figure 1 illustrates the Pearson correlation matrix of the original dataset, revealing strong positive correlations among zone-related variables, moderate relationships between temperature and airflow-related features, and predominantly negative correlations between humidity and both thermal and flow-related variables. These patterns highlight the inherent interdependencies within the raw data and underscore the importance of appropriate preprocessing and model design to effectively capture both direct and inverse relationships in subsequent predictive modeling stages.

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

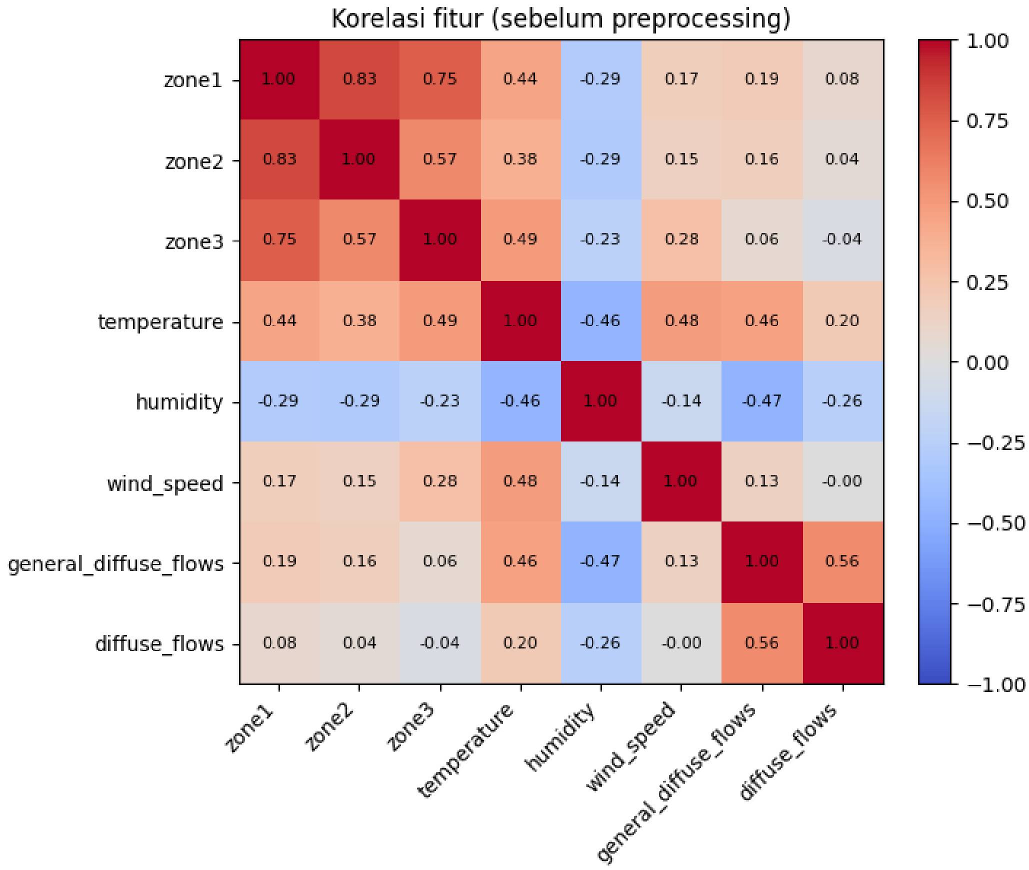

Figure 1: Correlation matrix of input features prior to preprocessing.

Temporal dependency analysis is a fundamental step in time-series modeling, as it enables the identification of inherent autocorrelation structures, seasonal patterns, and appropriate model configurations. To investigate the dynamic behavior of the aggregated energy consumption data recorded at 30-minute intervals, both the autocorrelation function (ACF) and partial autocorrelation function (PACF) were examined. As illustrated in Figure 2, the ACF plot exhibits a slowly decaying oscillatory pattern, indicating strong temporal persistence and potential periodicity within the series, while the PACF highlights significant spikes at early lags and around the daily cycle, suggesting the presence of both short-term dependencies and longer-range seasonal effects. These characteristics confirm the non-random nature of the energy consumption sequence and provide critical insights for selecting suitable lag structures and model architectures in subsequent forecasting stages.

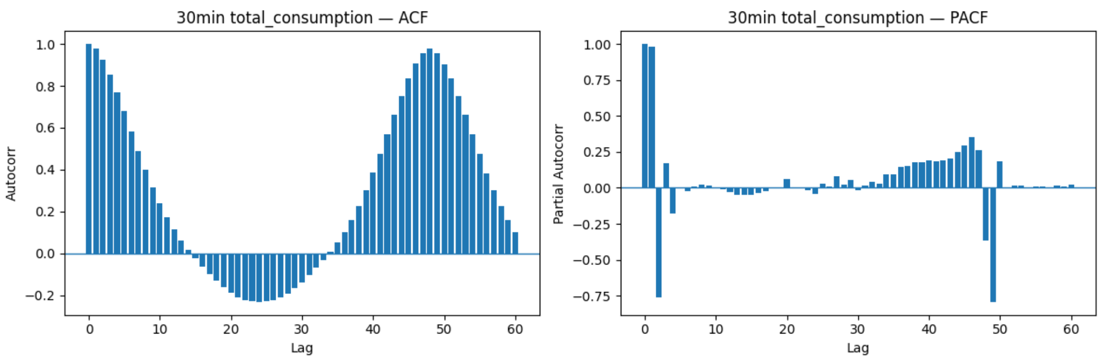

Figure 2: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the 30-minute total energy consumption time series.

Interpreting the contribution of individual input features is essential for enhancing the transparency, reliability,

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

and practical applicability of data-driven energy prediction models. To this end, a SHapley Additive exPlanations (SHAP) analysis was performed to quantify the marginal impact of each explanatory variable on the model output across the entire dataset. As shown in Figure 3, the SHAP summary plot ranks features according to their overall importance while simultaneously illustrating the direction and magnitude of their influence. Temporal variables, particularly hour and month, exhibit the strongest contributions, reflecting pronounced diurnal and seasonal consumption patterns. Meteorological variables such as temperature and airflow-related features also demonstrate notable effects, whereas categorical indicators such as weekend and day of week show comparatively lower influence. This analysis provides valuable insights into the internal decision-making process of the model and confirms that the learned relationships are physically and operationally consistent with real-world energy consumption behavior.

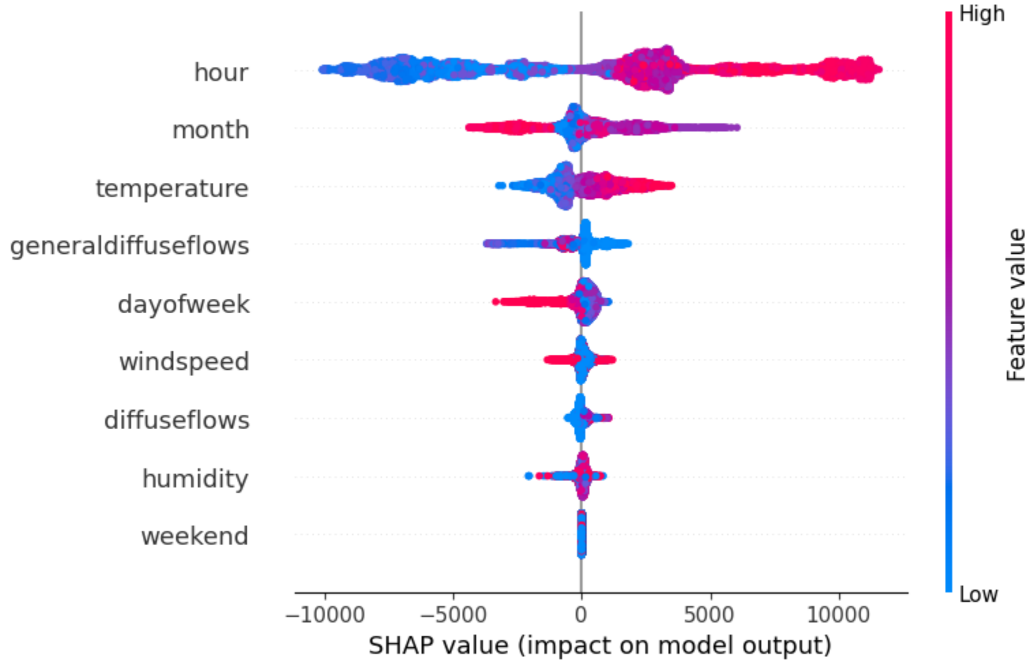

Figure 3: SHAP summary plot illustrating the impact of input features on the model output.

3.2 Data Preprocessing

Data preprocessing represents a fundamental component in the development of reliable and robust electricity consumption forecasting models, particularly when dealing with high-resolution SCADA-based time-series data. The dataset used in this study consists of heterogeneous variables measured at ten-minute intervals, including meteorological parameters, environmental indicators, and electricity consumption values across multiple distribution zones. In such a context, careful preprocessing is essential to ensure data quality, numerical stability, and effective learning by deep neural network architectures.

The preprocessing procedure began with a comprehensive assessment of data completeness and temporal consistency. All recorded variables were examined to identify potential missing values, anomalous entries, or discontinuities in the time series. Particular attention was given to the regularity of the ten-minute sampling intervals, as sequential learning models rely heavily on uninterrupted temporal sequences. When isolated missing observations were detected, appropriate corrective strategies were considered to preserve temporal continuity while minimizing the introduction of artificial distortions. In cases where longer gaps were identified, the affected segments were excluded from the analysis to maintain the integrity of the dataset. This verification process ensured that the final dataset used for model development was coherent and suitable for time-series forecasting applications.

Following data completeness verification, feature scaling and normalization were applied to address the heterogeneous numerical ranges of the input variables. The dataset includes features with substantially

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

different magnitudes, such as meteorological measurements and electricity consumption values expressed in different units and scales. Without appropriate normalization, features with larger numerical ranges may disproportionately influence the learning process, leading to biased parameter updates and unstable convergence behavior. To mitigate this issue, all continuous input variables were transformed using normalization techniques that map feature values onto a common numerical scale. This step improved numerical stability during training and facilitated efficient gradient-based optimization, which is particularly critical for deep learning models.

In addition to numerical preprocessing, temporal feature engineering was incorporated to enhance the representation of time-dependent consumption patterns. Electricity demand is inherently influenced by temporal factors such as daily activity cycles, weekly routines, and seasonal variations. To capture these effects explicitly, time-related attributes were derived from the original timestamp information, including indicators corresponding to the hour of the day, the day of the week, and the progression of the annual cycle. By integrating these temporal descriptors into the input space, the forecasting models are provided with structured contextual information that supports the learning of periodic and cyclical consumption behaviors characteristic of urban electricity demand.

The final stage of preprocessing involved the analysis of correlation and redundancy among input features. Meteorological and environmental variables often exhibit interdependencies due to shared physical drivers, which can result in redundant information within the feature set. High levels of redundancy may increase computational complexity and negatively affect model generalization by encouraging overfitting. To address this issue, statistical correlation analysis was conducted to quantify relationships among the input variables and identify strongly correlated feature pairs. The outcomes of this analysis informed subsequent optimization stages by highlighting candidates for feature reduction, thereby contributing to a more compact and informative representation of the input space prior to model training.

3.3 Deep Learning Models

Deep learning has emerged as a powerful modeling paradigm for complex energy-demand systems due to its ability to learn nonlinear and hierarchical representations directly from data. In short-term electricity consumption forecasting, particularly at high temporal resolutions, demand patterns are influenced by a combination of meteorological conditions, environmental factors, and human activity cycles, leading to strong temporal dependence and non-stationary behavior. These characteristics pose significant challenges for traditional statistical and shallow learning models, which often rely on restrictive assumptions and limited feature representations. Deep learning architectures, by contrast, can automatically extract informative latent features from multivariate inputs and model temporal dependencies through sequential learning mechanisms, making them well suited for high-frequency electricity load forecasting using SCADA data.

In this study, five deep learning architectures were selected to provide a comprehensive and systematic modeling framework for short-term electricity consumption prediction. These models include ANN, RNN, LSTM, GRU, and BiLSTM. The selected architectures represent a spectrum of modeling capabilities, ranging from static nonlinear approximation to advanced sequence learning with memory mechanisms. Their inclusion enables a structured comparison of how different architectural designs capture the temporal dynamics of electricity consumption recorded at ten-minute intervals across multiple urban distribution zones.

ANN was employed as a feedforward baseline model to establish a reference level of nonlinear predictive capability. ANN architectures consist of interconnected layers of neurons, where each neuron computes a weighted sum of its inputs followed by a nonlinear activation function. This structure allows ANN to approximate complex nonlinear mappings between explanatory variables, such as meteorological and environmental factors, and electricity consumption. However, ANN processes each observation independently and does not explicitly account for temporal correlations between successive time steps. As a result, while ANN can model instantaneous nonlinear relationships, its ability to capture sequential dependencies inherent in electricity consumption time series is limited.

To explicitly incorporate temporal dependence, RNN was introduced as a recurrent baseline model. RNN architectures include recurrent connections that propagate information from previous time steps to the

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

current hidden state, enabling the model to learn temporal relationships in sequential data. This internal memory mechanism makes RNN conceptually suitable for time-series forecasting tasks. In practice, however, conventional RNNs may suffer from training instability due to vanishing and exploding gradient problems, particularly when modeling long or high-frequency sequences. These limitations can restrict their effectiveness in capturing long-term dependencies in electricity consumption data.

LSTM was adopted to overcome the shortcomings of standard RNNs by introducing gated memory cells that regulate the flow of information through the network. The LSTM architecture employs input, forget, and output gates to selectively retain or discard information over time, thereby improving gradient flow and training stability. This capability enables LSTM models to capture both short-term fluctuations and longer-term temporal patterns in electricity consumption, which are influenced by daily routines, weekly cycles, and seasonal variations. Consequently, LSTM has been widely applied in short-term load forecasting and related energy prediction tasks.

GRU was also considered as a more compact gated recurrent architecture. GRU simplifies the LSTM design by merging gating mechanisms and eliminating the explicit memory cell, resulting in fewer trainable parameters and reduced computational complexity. Despite this simplification, GRU retains the ability to model temporal dependencies effectively. Including GRU in the modeling framework allows for an assessment of the trade-off between architectural complexity, training efficiency, and temporal learning capability in high-frequency electricity consumption forecasting.

BiLSTM was incorporated to further enhance temporal feature learning through bidirectional sequence processing. Unlike unidirectional recurrent models that process input sequences only in the forward temporal direction, BiLSTM employs two parallel recurrent layers that process the sequence in both forward and backward directions. The outputs of these layers are then combined to form a richer temporal representation. This bidirectional structure enables the model to leverage contextual information from both past and future observations during training, which can improve the learning of complex consumption trajectories and transition patterns. Due to its enhanced representational capacity, BiLSTM was selected as the core predictive architecture for subsequent optimization within the proposed framework.

3.4 Metaheuristic Optimization Algorithms

The predictive performance of deep learning models is strongly influenced by the selection of hyperparameters that govern both network structure and training dynamics. In short-term electricity consumption forecasting, inappropriate hyperparameter configurations may lead to slow convergence, unstable training behavior, or poor generalization to unseen time periods. Given the nonlinear and non-convex nature of the hyperparameter search space, manual tuning is inefficient and difficult to reproduce, while exhaustive search strategies rapidly become computationally prohibitive. These challenges motivate the adoption of metaheuristic optimization algorithms, which provide flexible and derivative-free search mechanisms suitable for complex black-box optimization problems.

Metaheuristic algorithms operate by iteratively improving a population of candidate solutions through stochastic update rules that balance global exploration and local exploitation. Such algorithms are particularly effective for hyperparameter optimization of deep learning models, where the objective function is typically evaluated through validation-based forecasting error and does not possess an explicit analytical form. In the context of high-resolution electricity load forecasting, metaheuristic optimization supports the identification of robust hyperparameter configurations that enhance training stability and generalization under realistic operating conditions.

In this study, metaheuristic optimization is employed to automate hyperparameter tuning for the deep learning models, with a primary focus on BiLSTM as the core forecasting architecture. The optimization problem is formulated such that each candidate solution encodes a feasible hyperparameter vector, while its fitness is evaluated using a predefined objective function computed on a validation subset obtained through a time-series–consistent data split. This formulation ensures that optimization targets generalization capability rather than memorization of training data.

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

The Ninja Optimization Algorithm (NiOA) is adopted as the principal optimization engine within the proposed framework. NiOA is a population-based metaheuristic that emphasizes adaptive exploration and exploitation strategies to navigate complex optimization landscapes. During the exploration phase, candidate solutions traverse the search space broadly to promote diversity and reduce sensitivity to initial conditions. Exploitation mechanisms are subsequently applied to refine promising solutions and accelerate convergence toward optimal configurations. Mutation and adaptive update strategies are integrated to prevent stagnation and enhance robustness against local optima.

In addition to NiOA, several state-of-the-art metaheuristic algorithms are incorporated to provide a comprehensive comparative optimization framework. These include WAO, BBO, GA, SFS, DE, and JAYA. Each of these algorithms embodies a distinct search philosophy and population update mechanism, enabling diverse exploration–exploitation behaviors within the hyperparameter space. By evaluating these algorithms under consistent experimental conditions, the study enables a systematic assessment of different metaheuristic strategies for deep learning hyperparameter optimization in short-term electricity consumption forecasting.

3.5 Proposed NiOA-Based Optimization Framework

The Ninja Optimization Algorithm (NiOA) is adopted in this study as the principal metaheuristic framework for automated hyperparameter optimization of the BiLSTM-based electricity consumption forecasting model. NiOA is a population-based optimization algorithm designed to efficiently navigate complex, nonlinear, and high-dimensional search spaces by adaptively balancing global exploration and local exploitation. Such characteristics make NiOA particularly suitable for deep learning hyperparameter optimization, where the objective function is typically non-convex, computationally expensive to evaluate, and available only in a black-box form through validation-based forecasting performance.

Within the proposed framework, the hyperparameter optimization task is formulated as a continuous optimization problem. Each ninja agent encodes a candidate hyperparameter configuration for the BiLSTM model as a real-valued vector within predefined feasibility bounds. Candidate quality is evaluated using a fitness function computed on a validation subset obtained via a time-series–consistent split, ensuring that the optimization process targets generalization capability under realistic forecasting conditions.

NiOA proceeds through exploration, oscillatory diversification, mutation, and exploitation/refinement mechanisms. During exploration, the position of a search agent L_s at iteration t is updated according to:

$$L_s(t+1) = \begin{cases} L_s(t) + r_1 \cdot \left(L_s(t_1) - L_s(t_2)\right), & \text{if the exploration condition holds,} \\ \text{Random } L_s \in FS, & \text{otherwise,} \end{cases} \tag{1}$$

where $r_1 \in [0, 1]$ is a random scaling factor, t_1 and t_2 are randomly selected previous iterations, and FS denotes the feasible hyperparameter search space. This update promotes global exploration by encouraging candidate solutions to traverse diverse regions of the search landscape.

To further enhance diversification and mitigate premature convergence, an oscillatory movement component is applied:

$$D_s(t+1) = D_s(t) + |D_s(t) + r_2 \cdot D_s(t)| \cdot \cos(2\pi t), \tag{2}$$

where $r_2 \in [0, 1]$ introduces stochastic variation. The cosine term periodically alters the movement direction and magnitude, enabling broader coverage of the search space.

In addition, NiOA integrates a mutation mechanism to increase population diversity and facilitate escaping local optima. The mutation operator is defined as:

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

$$N = \sum_{n=0}^{a-1} \frac{(-1)^n}{2n+1} \cdot x^{(2n+1)}, \tag{3}$$

where a is a randomly generated integer controlling mutation intensity. This operator perturbs candidate solutions in a controlled manner, improving the algorithm's ability to explore alternative configurations.

Once promising regions are identified, NiOA intensifies local refinement through exploitation, governed by:

$$M_s(t+1) = J_1 M_s(t) + 2J_2 \cdot (M_s(t) + (M_s(t) + J_1)) \left(1 - \frac{M_s(t)}{M_s(t) + J_1}\right)^2, \tag{4}$$

where J_1 and J_2 are adaptive control parameters regulating exploitation strength. This update refines solutions around promising areas to enhance convergence toward high-quality hyperparameter configurations.

To prevent stagnation, an adaptive update strategy is applied when fitness improvement is not observed over consecutive iterations. The best-solution update is expressed as:

$$B_s(t+1) = L_s(t+1) + i \cdot n \cdot (L_s(t+1) - D_s(t+1)) + i \cdot n \cdot (M_s(t+1) + 2v_s \cdot R_s(t+1)), \tag{5}$$

where i, n, and v_s regulate update intensity, and $R_s(t+1)$ denotes the resource-related vector used in the NiOA update logic. This mechanism dynamically perturbs the current best solution to restore exploration when the search becomes trapped around suboptimal regions.

Algorithm 1 summarizes the overall workflow of the proposed NiOA-based hyperparameter optimization framework, explicitly linking the algorithmic steps to the update rules in Eqs. (1)–(5). By integrating NiOA with BiLSTM, the proposed framework provides a systematic and reproducible approach for automated hyperparameter optimization in short-term electricity consumption forecasting using high-resolution SCADA data [30].

3.6 State-of-the-Art Metaheuristic Models

In addition to NiOA, several state-of-the-art metaheuristic optimization algorithms are incorporated to establish a comparative framework for deep learning hyperparameter tuning under consistent experimental conditions. The inclusion of multiple metaheuristics enables a systematic evaluation of different exploration–exploitation philosophies, population update strategies, and convergence behaviors when applied to the BiLSTM-based short-term electricity consumption forecasting problem.

WAO is incorporated as a representative swarm-based metaheuristic with mechanisms that balance global exploration and local exploitation through iterative position updates guided by solution quality. This balance supports efficient navigation of complex search landscapes while maintaining sufficient diversity to reduce the risk of premature convergence.

BBO is considered as a population-based approach that models candidate solutions as habitats and employs information-sharing dynamics through migration mechanisms. Such structured exchange can help propagate high-quality solution components across the population while preserving diversity.

GA is included as a classical evolutionary benchmark. GA evolves candidate hyperparameter configurations using selection, crossover, and mutation operators, enabling robust global search in large and nonlinear spaces.

SFS is incorporated as a stochastic diffusion-based method that integrates exploration through random walks with localized refinement. This dual behavior is beneficial for escaping local optima and exploring high-dimensional hyperparameter spaces.

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Algorithm 1 NiOA-based hyperparameter optimization for BiLSTM

Input: Feasible search space FS, population size N, maximum iterations T, stagnation threshold τ
Input: Random coefficients $r_1, r_2 \in [0, 1]$, control parameters J_1, J_2, and factors i, n, v_s
Output: Best hyperparameter vector B
Initialize a population of N agents $\{L_s(0)\}_{s=1}^{N}$ uniformly in FS
Initialize auxiliary states $\{D_s(0)\}_{s=1}^{N}$, $\{M_s(0)\}_{s=1}^{N}$, and $\{R_s(0)\}_{s=1}^{N}$
Evaluate fitness of each agent by training BiLSTM using the encoded hyperparameters and computing validation fitness
Set best solution $B \leftarrow \arg\min_s$ fitness$(L_s(0))$ (or $\arg\max$ depending on the fitness definition)
Set stagnation counter $c \leftarrow 0$
for $t = 0$ to $T - 1$ **do**
    for each agent $s = 1, 2, \ldots, N$ **do**
        Exploration: update $L_s(t + 1)$ according to Eq. (1)
        Oscillatory diversification: update $D_s(t + 1)$ using Eq. (2)
        Mutation: compute mutation operator N using Eq. (3) and perturb the candidate if applicable
        Exploitation: refine $M_s(t + 1)$ according to Eq. (4)
        Enforce feasibility constraints (project into FS if needed)
        Evaluate fitness of the updated candidate using the validation set
    end for
    Update current best B_{new} from the population based on fitness
    if fitness(B_{new}) improves over fitness(B) **then**
        $B \leftarrow B_{\text{new}}$; $c \leftarrow 0$
    else
        $c \leftarrow c + 1$
    end if
    if $c \geq \tau$ **then**
        Stagnation handling: update the best solution using Eq. (5)
        Reset $c \leftarrow 0$
    end if
end for
Return B

DE is included as a population-based continuous optimizer that perturbs candidate solutions using scaled vector differences among randomly selected individuals, enabling efficient exploration of numerical hyperparameters and competitive convergence behavior in continuous domains.

JAYA is employed as a parameter-free algorithm that iteratively moves solutions toward the current best and away from the current worst candidate, reducing the need for algorithm-specific tuning while maintaining competitive search capability.

By benchmarking NiOA against WAO, BBO, GA, SFS, DE, and JAYA, the study provides a robust basis for comparing optimization behaviors and identifying effective strategies for deep learning hyperparameter tuning in high-resolution urban electricity consumption forecasting.

3.7 Evaluation Metrics

A rigorous and comprehensive evaluation framework is essential for assessing the predictive capability, robustness, and generalization performance of deep learning models developed for short-term electricity consumption forecasting. Given the complexity of urban electricity distribution systems and the nonlinear, time-dependent nature of electricity demand, reliance on a single evaluation metric may lead to incomplete or misleading conclusions. Therefore, this study employs a diverse set of regression-based performance metrics that collectively capture error magnitude, systematic bias, correlation strength, relative performance, and overall agreement between predicted and observed electricity consumption values.

The selected metrics include Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R^2),

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Relative Root Mean Squared Error (RRMSE), Nash–Sutcliffe Efficiency (NSE), and the Willmott Index of Agreement (WI). These indicators are widely adopted in electricity load forecasting, energy systems modeling, and power system analysis, enabling meaningful comparison with related studies and ensuring methodological consistency.

Error-based metrics form the foundation of predictive performance assessment. MSE quantifies the average squared deviation between predicted and observed electricity consumption values and penalizes large prediction errors more heavily, making it sensitive to extreme deviations. RMSE, derived from MSE, retains the same physical unit as the target variable and is therefore particularly useful for interpreting prediction accuracy in practical electricity consumption terms. MAE provides a linear measure of average absolute deviation and offers a more robust assessment in the presence of outliers. Together, these metrics provide complementary perspectives on overall prediction error.

MBE is employed to quantify systematic bias in model predictions. Unlike absolute error metrics, MBE reveals whether a model exhibits a consistent tendency to overestimate or underestimate electricity consumption. Bias assessment is especially important in electricity distribution applications, as persistent overprediction or underprediction may lead to inefficient operational planning, load imbalance, and suboptimal grid management decisions.

To evaluate the strength of association between predicted and observed values, correlation-based metrics are incorporated. The Pearson correlation coefficient (r) measures the degree of linear relationship between predictions and observations, providing insight into how well the model captures temporal variability and load dynamics. The coefficient of determination (R^2) quantifies the proportion of variance in observed electricity consumption explained by the model, serving as an indicator of explanatory power and goodness of fit.

Relative performance metrics are also considered to enable scale-independent assessment. RRMSE normalizes RMSE with respect to the mean of the observed electricity consumption values, facilitating comparison across datasets with different magnitudes, seasonal conditions, or spatial zones. Such normalization is particularly valuable in electricity demand studies where consumption levels may vary substantially over time and across distribution areas.

Efficiency- and agreement-based metrics are included to provide a holistic evaluation of model performance. NSE compares model predictions against the mean of observed values and assesses how well the model reproduces observed electricity demand dynamics relative to a baseline predictor. WI offers a bounded measure of agreement that accounts for both systematic and unsystematic errors, making it a robust indicator of overall predictive reliability in time-series electricity forecasting applications.

Table 1 summarizes the mathematical definitions of all evaluation metrics employed in this study.

In the above equations, y_i and $\hat{y}_i$ denote the observed and predicted electricity consumption values at the i-th time step, respectively, $\bar{y}$ and $\bar{\hat{y}}$ represent their corresponding mean values, and N denotes the total number of samples. The combined use of these complementary metrics enables a robust and multidimensional evaluation of model performance, ensuring that predictive accuracy, bias characteristics, correlation structure, and overall agreement are thoroughly assessed.

4 Experimental Results

4.1 Baseline Model Performance (Before Optimization)

This subsection presents a detailed quantitative and qualitative comparison of the baseline deep learning models applied to short-term electricity consumption forecasting prior to optimization. The evaluated models include ANN, RNN, LSTM, GRU, and BiLSTM. The primary objectives of this analysis are to establish a robust performance benchmark using multiple regression metrics and to identify architectural and learning limitations that motivate the subsequent application of metaheuristic optimization techniques. The numerical results corresponding to all baseline models are summarized in Table 2.

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Table 1: Regression-based evaluation metrics used for model performance assessment

Metric	Mathematical Definition
Mean Squared Error (MSE)	$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i - \hat{y}_i\rvert$
Mean Bias Error (MBE)	$\text{MBE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)$
Pearson Correlation Coefficient (r)	$r = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}$
Coefficient of Determination (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Relative RMSE (RRMSE)	$\text{RRMSE} = \frac{\text{RMSE}}{\bar{y}} \times 100$
Nash–Sutcliffe Efficiency (NSE)	$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$
Willmott Index (WI)	$\text{WI} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(\lvert\hat{y}_i - \bar{y}\rvert + \lvert y_i - \bar{y}\rvert)^2}$

Table 2: Performance comparison of baseline deep learning models before optimization

Model	MSE	RMSE	MAE	MBE	r	R^2	RRMSE	NSE	WI
BiLSTM	0.0093	0.0964	0.0598	0.0395	0.876	0.854	1.69	0.892	0.891
GRU	0.0134	0.1158	0.0679	0.0510	0.867	0.846	1.97	0.878	0.875
LSTM	0.0192	0.1385	0.0765	0.0589	0.856	0.836	2.29	0.864	0.864
RNN	0.0264	0.1624	0.0881	0.0738	0.844	0.823	2.63	0.847	0.852
ANN	0.0351	0.1874	0.1021	0.0864	0.834	0.813	3.01	0.832	0.838

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Table 2 reveals a clear performance hierarchy that is consistent across error-based, correlation-based, and agreement-based metrics. In general, architectures that explicitly model temporal dependence exhibit superior predictive behavior compared to static feedforward learning. This observation is particularly expected in the present forecasting task because the target variable represents electricity consumption measured at a high temporal resolution (10-minute intervals), where short-term variability is strongly driven by temporal continuity, regime transitions, and time-of-day activity patterns. Under such conditions, capturing sequential dependencies becomes essential for learning the intrinsic load dynamics embedded in SCADA time series.

The ANN model exhibits the weakest predictive performance among all evaluated approaches. This behavior is reflected by its comparatively high error values, including an MSE of 0.0351, RMSE of 0.1874, and MAE of 0.1021. Furthermore, ANN presents the largest systematic bias (MBE = 0.0864), indicating a consistent deviation between predicted and observed electricity consumption values. The relatively low correlation coefficient ($r = 0.834$) and coefficient of determination ($R^2 = 0.813$) confirm the limited ability of the feedforward architecture to capture the temporal dynamics inherent in high-resolution electricity consumption time series. Conceptually, this limitation arises because ANN treats each time step as an independent sample and therefore cannot exploit the sequential continuity and temporal persistence that characterize load trajectories. As a consequence, ANN is less capable of learning the demand evolution patterns that emerge from repeated daily cycles, weekend–weekday differences, and short-term fluctuations that occur within a day at ten-minute granularity.

The RNN model demonstrates improved performance relative to ANN, highlighting the benefit of incorporating recurrent connections for temporal learning. The RNN achieves reduced error metrics (MSE = 0.0264, RMSE = 0.1624, MAE = 0.0881) and improved correlation ($r = 0.844$; $R^2 = 0.823$). However, the model still exhibits notable bias (MBE = 0.0738) and elevated relative error (RRMSE = 2.63), indicating that conventional recurrent structures remain limited in capturing longer-term dependencies in electricity demand. The observed gap between RNN and gated architectures is consistent with the known tendency of vanilla recurrent networks to struggle with stable memory retention across longer sequences, especially when the data exhibit mixed short-term volatility and longer periodic components. In high-frequency SCADA series, short-term transitions can be abrupt, and the ability to preserve relevant context over time becomes increasingly important for robust forecasting.

The introduction of gated recurrent architectures leads to further performance gains. The LSTM model significantly reduces prediction error (RMSE = 0.1385; MAE = 0.0765) while improving correlation and explanatory power ($r = 0.856$; $R^2 = 0.836$). These improvements can be attributed to the gated memory structure of LSTM, which enables more effective retention and selective updating of temporal information over extended horizons. In the present application, this capability is particularly valuable because electricity demand at a given ten-minute interval may depend simultaneously on very recent demand levels (capturing short-term autocorrelation) and on broader contextual patterns such as diurnal activity and seasonal influences. The reduction in MBE (0.0589) relative to ANN and RNN suggests that incorporating gating not only improves accuracy but also reduces systematic deviation, which is important for operational planning applications where persistent overestimation or underestimation may degrade decision quality.

The GRU model further enhances predictive performance while maintaining reduced architectural complexity. GRU achieves RMSE = 0.1158 and MAE = 0.0679, together with improved agreement and efficiency metrics (NSE = 0.878; WI = 0.875). These results indicate that GRU provides a favorable balance between model complexity and temporal learning capability in short-term electricity consumption forecasting. The improvements relative to LSTM in Table 2 suggest that the simplified gating structure of GRU can be advantageous when the dataset exhibits strong short-horizon dependencies and frequent transitions, as is typical in high-frequency load records. Moreover, the reduced parameterization of GRU can support more stable training dynamics and lower risk of overfitting in scenarios where the temporal resolution is high and the effective sequence length becomes large.

Among all baseline models, BiLSTM delivers the strongest predictive performance prior to optimization. The BiLSTM model achieves the lowest error values (MSE = 0.0093; RMSE = 0.0964; MAE = 0.0598), the smallest bias (MBE = 0.0395), and the highest association and agreement metrics ($r = 0.876$; $R^2 = 0.854$; NSE = 0.892; WI = 0.891). These gains are attributed to the bidirectional learning mechanism, which allows the model to exploit richer contextual representations during training by processing sequences in forward and backward directions. Although operational forecasting in real time relies on past information, bidirectional training can still yield stronger internal representations and improved generalization because it enables the network to learn more complete temporal structures from historical sequences. In SCADA-based datasets that span an entire year, this can help capture recurring patterns, transitions across seasons, and changes in consumption regimes with higher fidelity.

Despite the superior baseline performance of BiLSTM, Table 2 indicates the presence of non-negligible residual errors and bias. This suggests that baseline deep learning models remain sensitive to hyperparameter selection and may be affected by redundancy and correlation among explanatory variables. Since the dataset includes multiple meteorological and environmental drivers that may exhibit partial overlap in information content, the learning process can be influenced by feature interactions and scaling effects. These observations motivate the integration of metaheuristic optimization to systematically refine model configurations and further enhance forecasting robustness and generalization in high-frequency urban electricity consumption prediction tasks.

A comprehensive evaluation of predictive model performance requires not only point estimates of accuracy metrics but also an assessment of their variability and robustness across different modeling approaches. To this end, a distributional analysis of multiple statistical performance indicators was conducted to facilitate a holistic comparison among the evaluated models. Figure 4 presents box plots with horizontal swarm overlays for a diverse set of error-based and goodness-of-fit metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott Index (WI). The visualized distributions provide insights into central tendency, dispersion, and the presence of outliers, thereby enabling a nuanced comparison of model stability and predictive reliability. Such an analysis is particularly important in energy consumption forecasting, where consistent performance across varying conditions is as critical as achieving low average error.

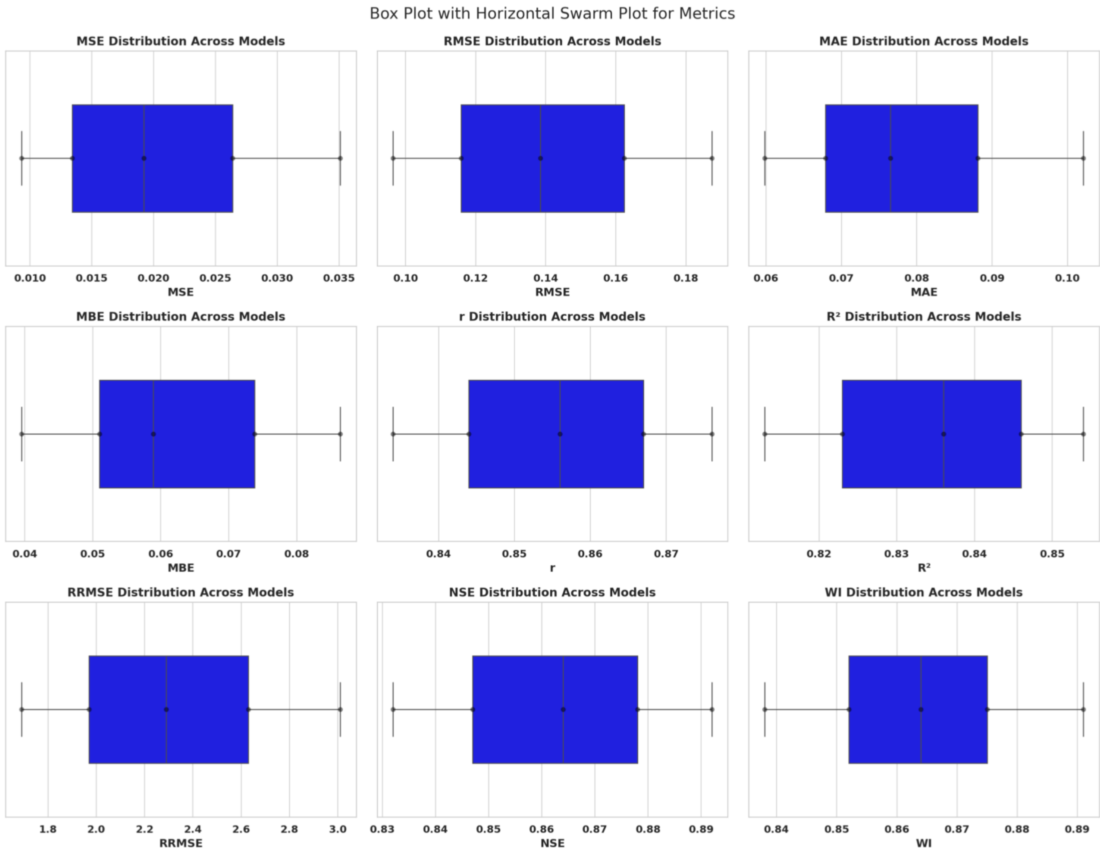

Figure 4: Box plots with horizontal swarm distributions of performance metrics across the evaluated models.

Assessing the distributional characteristics of model performance metrics is an important step in validating the robustness and statistical reliability of predictive models. In particular, examining whether evaluation metrics conform to theoretical distributions provides insight into the presence of skewness, outliers, or deviations from

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

normality that may affect comparative analysis and inferential conclusions. To this end, quantile–quantile (Q–Q) plots were generated for all considered performance indicators. As illustrated in Figure 5, the empirical quantiles of metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott Index (WI) are plotted against their corresponding theoretical quantiles. The close alignment of the data points with the reference lines suggests that the distributions of these metrics exhibit near-normal behavior, thereby supporting the validity of subsequent statistical comparisons across models.

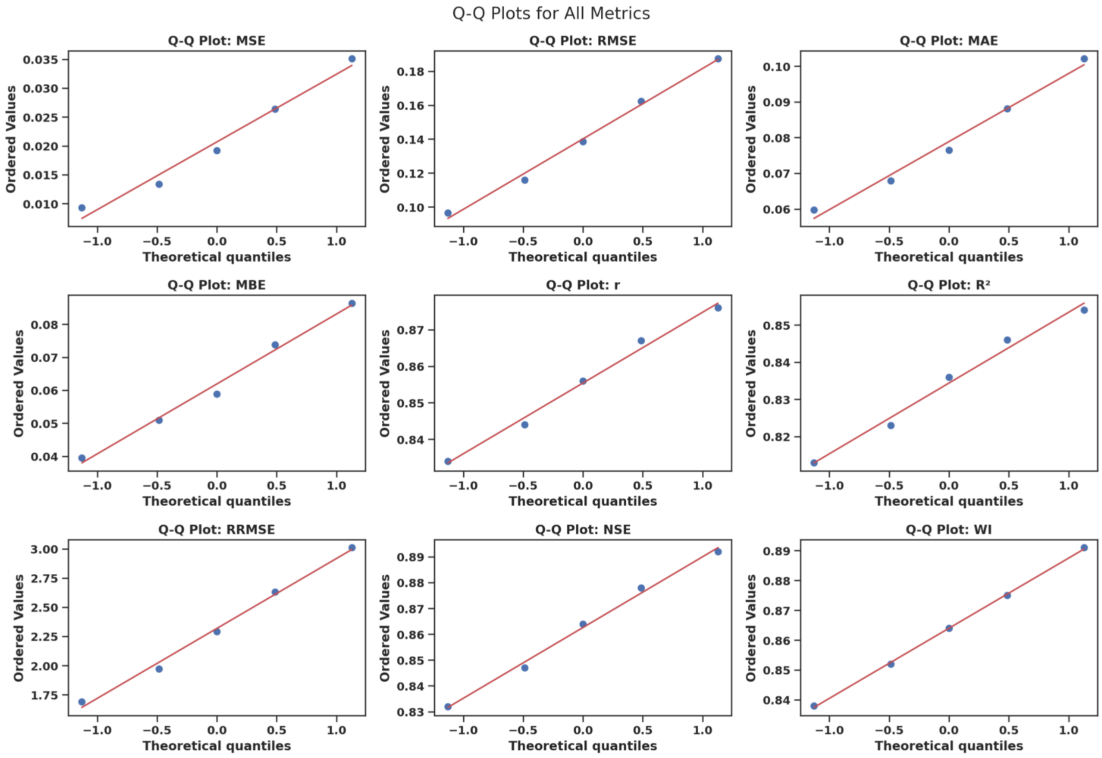

Figure 5: Quantile–quantile (Q–Q) plots of performance metrics across the evaluated models.

Beyond median- and distribution-based comparisons, incorporating measures of central tendency and dispersion offers a more complete understanding of model performance consistency. In this study, the mean and standard deviation of each evaluation metric were explicitly visualized to assess both average predictive accuracy and variability across the examined models. As illustrated in Figure 6, box plots augmented with reference lines representing the mean as well as one standard deviation above and below the mean are presented for a comprehensive set of performance indicators, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott Index (WI). This combined visualization facilitates a simultaneous evaluation of accuracy, robustness, and dispersion, thereby enabling a clearer identification of models that achieve not only low error but also stable performance across varying conditions.

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

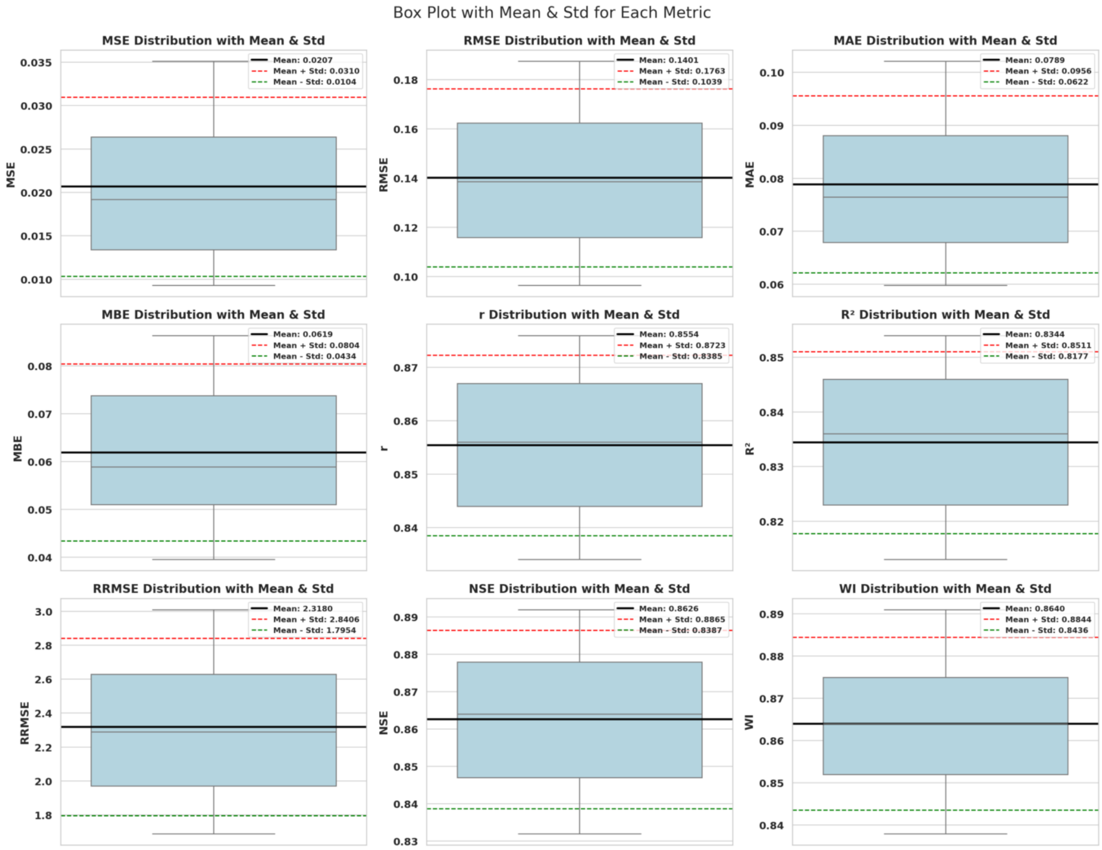

Figure 6: Box plots of performance metrics with mean and standard deviation indicators across the evaluated models.

4.2 Optimized Model Analysis

This subsection provides a comprehensive analysis of the optimized hybrid models obtained by integrating BiLSTM with different metaheuristic optimization algorithms. The objective is to assess how various optimization strategies influence the predictive behavior of BiLSTM under a unified evaluation framework. The optimized hybrid models considered include NijOA+BiLSTM, WAO+BiLSTM, BBO+BiLSTM, GA+BiLSTM, SFS+BiLSTM, DE+BiLSTM, and JAYA+BiLSTM. The corresponding quantitative results are summarized in Table 3.

Table 3: Performance analysis of optimized hybrid models based on BiLSTM

Model	**MSE**	**RMSE**	**MAE**	**MBE**	r	R^2	**RRMSE**	**NSE**	**WI**
NijOA + BiLSTM	1.45E-05	0.0038	0.00019	2.60E-05	0.973	0.970	0.092	0.973	0.978
WAO + BiLSTM	9.80E-05	0.0099	0.00043	-9.00E-05	0.958	0.955	0.170	0.964	0.967
BBO + BiLSTM	0.000125	0.0112	0.00047	0.00011	0.951	0.949	0.210	0.960	0.962
GA + BiLSTM	0.000190	0.0138	0.00051	0.00015	0.939	0.936	0.330	0.953	0.952
SFS + BiLSTM	0.000230	0.0151	0.00054	-0.00017	0.931	0.929	0.420	0.948	0.946
DE + BiLSTM	0.000290	0.0170	0.00056	0.00019	0.925	0.921	0.570	0.943	0.940
JAYA + BiLSTM	0.000380	0.0195	0.00057	-0.00021	0.917	0.913	0.810	0.938	0.934

Table 3 indicates that integrating metaheuristic optimization with BiLSTM yields consistently strong predictive performance across all evaluation criteria. Compared with the baseline results in Table 2, the optimized hybrid configurations exhibit markedly reduced error magnitudes, substantially improved correlation, and

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

higher agreement and efficiency indices. This global improvement pattern supports the central methodological premise of this study: deep learning performance in high-frequency electricity demand forecasting is highly sensitive to hyperparameter configuration, and systematic search-based tuning provides an effective mechanism for enhancing both accuracy and robustness.

Across the optimized set, the correlation coefficient remains consistently high (from $r = 0.917$ to $r = 0.973$) and the coefficient of determination remains strong (from $R^2 = 0.913$ to $R^2 = 0.970$), confirming that the optimized models capture the temporal variability and demand dynamics embedded in SCADA time series. The agreement-based metrics (NSE and WI) remain close to unity across all optimizers, indicating that the hybrid models reproduce the observed demand trajectories with high fidelity. In practical forecasting contexts, such agreement is particularly important because operational planning decisions depend not only on pointwise error reduction but also on preserving the shape and temporal evolution of the demand signal.

Among all strategies, NijOA+BiLSTM achieves the most favorable results across all evaluation metrics. This dominant performance suggests that NijOA provides an effective balance between exploration and exploitation in the hyperparameter search landscape, enabling the discovery of highly competitive configurations for BiLSTM training. Importantly, the superiority of NijOA+BiLSTM is reflected consistently across error-based measures (MSE, RMSE, MAE), association measures (r, R^2), and agreement measures (NSE, WI), rather than being confined to a single criterion. Such consistency indicates that the gains are not merely numerical artifacts of a particular metric but reflect a broad and stable improvement in predictive behavior.

The second tier of performance is formed by WAO+BiLSTM and BBO+BiLSTM, which exhibit strong results across all metrics and remain close to the best-performing hybrid configuration. Their competitive behavior suggests that these optimizers also provide effective hyperparameter refinement strategies, albeit with slightly less consistent convergence to the globally most favorable region of the search space. These differences may be explained by distinct population update mechanisms and exploitation intensities, which can lead to different convergence trajectories in complex optimization landscapes.

The remaining optimizers (GA, SFS, DE, and JAYA) also yield strong predictive quality, though they exhibit a gradual reduction in performance relative to NijOA-, WAO-, and BBO-based tuning. Nevertheless, the fact that all optimizers improve BiLSTM performance highlights the general effectiveness of metaheuristic-guided hyperparameter tuning for high-frequency electricity consumption forecasting. In particular, the consistent reduction of error-based metrics combined with the preservation of high agreement indices suggests that optimization improves not only point accuracy but also generalization to unseen temporal segments, which is essential for operational forecasting in real distribution networks.

The bias behavior, as measured by MBE in Table 3, further provides insight into systematic prediction tendencies. The optimized models exhibit MBE values that remain very close to zero with varying sign, indicating that the optimization process also contributes to reducing persistent overestimation or underestimation tendencies. From an operational viewpoint, bias reduction is critical because systematic forecasting bias may lead to inefficient scheduling of distribution resources, inaccurate operational reserves, and avoidable stress on distribution assets.

Overall, the results in Table 3 establish that hybrid metaheuristic–BiLSTM forecasting provides a highly effective and robust approach for short-term electricity consumption prediction in an urban distribution setting. The systematic ranking of hybrid models across multiple metrics provides a strong empirical basis for selecting NijOA+BiLSTM as the most favorable configuration within the considered optimization pool, while also confirming that alternative optimizers offer competitive improvements and may be considered depending on computational constraints and deployment requirements.

When multiple predictive models are evaluated across a diverse set of performance indicators, a unified multi-dimensional visualization is valuable for capturing trade-offs and overall comparative behavior. To this end, a radar chart was employed to simultaneously assess the normalized performance of hybrid metaheuristic–Bidirectional Long Short-Term Memory (BiLSTM) models across several complementary metrics. As shown in Figure 7, the models optimized using different metaheuristic algorithms—namely Ninja Optimization Algorithm (NijOA), Whale Optimization Algorithm (WAO), Biogeography-Based Optimization (BBO), Genetic Algorithm (GA), Stochastic Fractal Search (SFS), Differential Evolution (DE), and JAYA—are compared in terms of error-based metrics (MSE, RMSE, MAE, MBE, and RRMSE) and goodness-of-fit indicators (r and R^2). This holistic representation enables an intuitive assessment of model strengths

DOI: https://doi.org/10.54216/JAIM.110101

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

and weaknesses, highlighting the balance between accuracy and consistency achieved by each hybrid configuration.

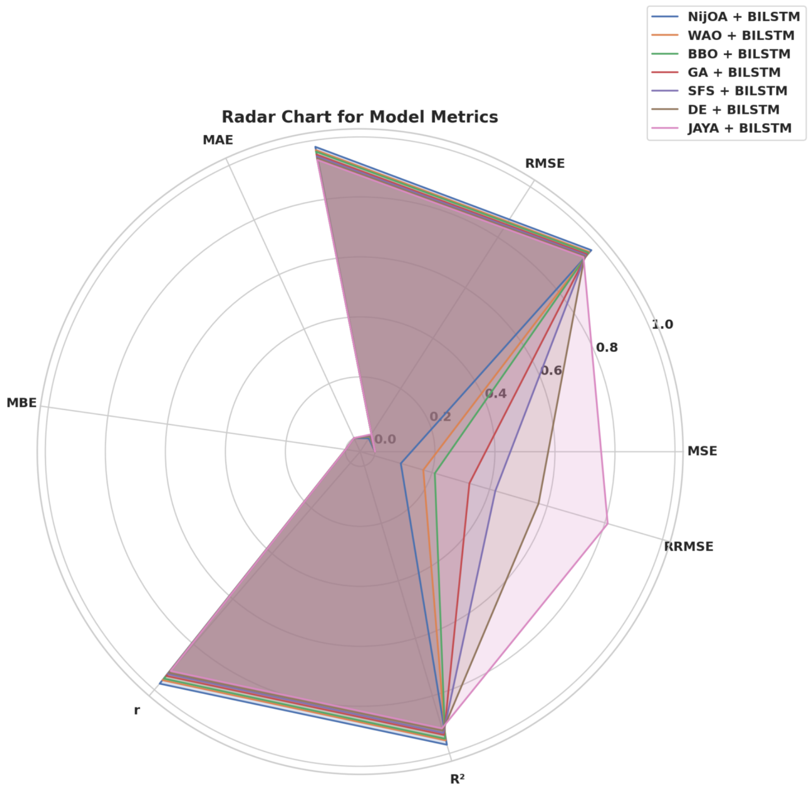

Figure 7: Radar chart comparing normalized performance metrics of hybrid metaheuristic–BiLSTM models.

Comparing predictive models across multiple, often competing, evaluation criteria requires visualization techniques capable of preserving high-dimensional performance information. Parallel coordinates plots are particularly effective in this context, as they allow each model to be simultaneously assessed across a set of normalized metrics while revealing relative strengths, weaknesses, and trade-offs. As illustrated in Figure 8, the performance of hybrid metaheuristic–Bidirectional Long Short-Term Memory (BiLSTM) models optimized using different algorithms is compared across error-based indicators (MSE, RMSE, MAE, MBE, and RRMSE) and goodness-of-fit metrics (r, R^2, Nash–Sutcliffe Efficiency (NSE), and Willmott Index (WI)). The convergence and divergence of the polylines across axes provide an intuitive representation of model dominance and balance, facilitating informed model selection based on overall predictive reliability rather than isolated metrics.

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

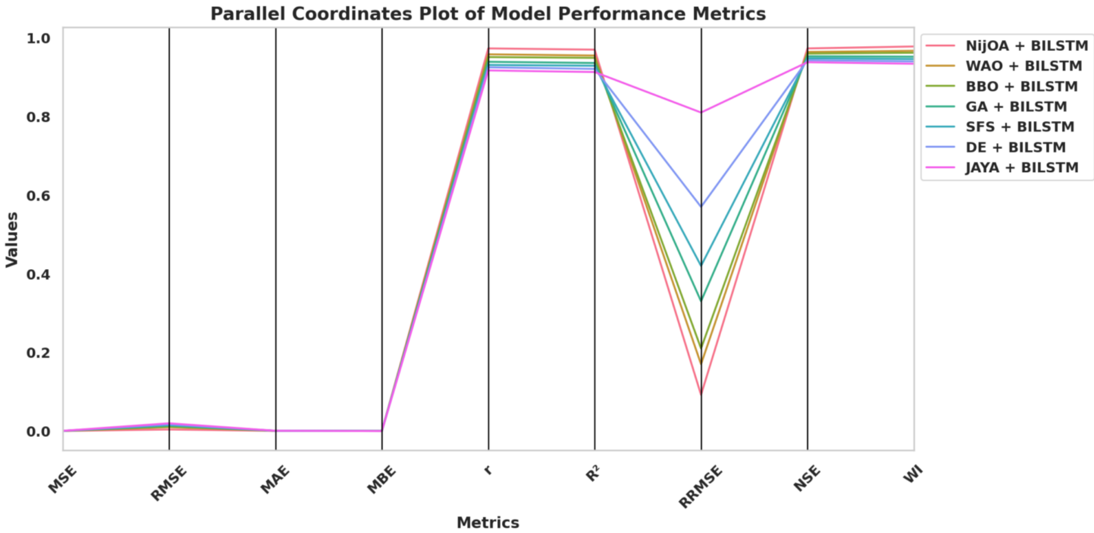

Figure 8: Parallel coordinates plot of normalized performance metrics for hybrid metaheuristic–BiLSTM models.

To facilitate a detailed, metric-wise comparison of the evaluated predictive models, a facet grid visualization was employed to disaggregate performance indicators into individual subplots. This approach enables a clearer inspection of how each model performs with respect to specific error-based and goodness-of-fit metrics without the visual overlap inherent in aggregated plots. As illustrated in Figure 9, the facet grid presents bar charts for Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), Pearson correlation coefficient (r), coefficient of determination (R^2), Relative RMSE (RRMSE), Nash–Sutcliffe Efficiency (NSE), and Willmott Index (WI). By isolating each metric, this visualization highlights relative ranking patterns and performance trade-offs across models, thereby supporting a more granular interpretation of predictive accuracy, bias, and goodness of fit.

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

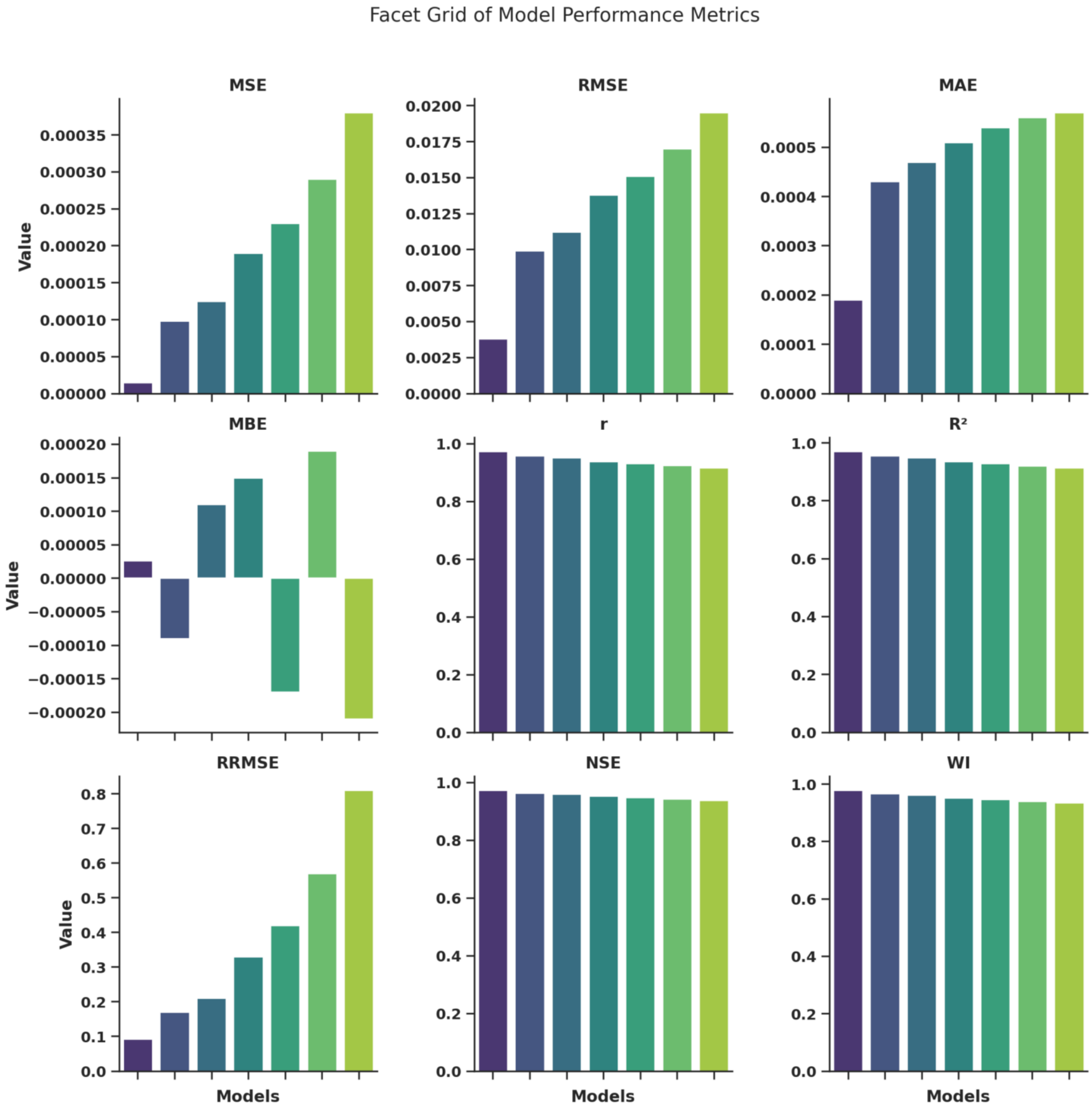

Figure 9: Facet grid visualization of performance metrics across the evaluated models.

5 Conclusion and Future Work

This study proposed and validated a comprehensive deep learning and metaheuristic optimization framework for short-term electricity consumption forecasting using high-resolution SCADA data collected from the urban distribution network of Tetouan city. By leveraging ten-minute interval measurements that capture fine-grained demand dynamics across multiple distribution zones, the proposed framework addresses key challenges associated with nonlinear behavior, temporal dependency, and sensitivity to model configuration in urban electricity demand forecasting.

The comparative analysis of baseline deep learning models confirms that architectures incorporating explicit temporal learning mechanisms consistently outperform feedforward structures when applied to high-frequency electricity consumption data. In particular, recurrent and gated models demonstrate superior capability in capturing the inherent temporal correlations and demand fluctuations observed in urban electricity loads. Among these architectures, BiLSTM exhibits the strongest baseline performance, which can be attributed to its bidirectional temporal representation that enables the model to exploit contextual information from both past and future time steps during training. This characteristic proves especially beneficial for modeling complex and rapidly varying electricity consumption patterns observed in SCADA-based datasets.

DOI: https://doi.org/10.54216/JAIM.110101
Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

Building upon the baseline analysis, the integration of metaheuristic optimization with the BiLSTM model leads to substantial and systematic performance enhancements across all evaluation metrics. The optimized hybrid models consistently demonstrate reduced prediction error, improved correlation with observed demand, and enhanced agreement and efficiency measures. Among the considered optimization strategies, the NijOA+BiLSTM configuration delivers the most accurate and robust forecasting performance. This outcome highlights the effectiveness of the Ninja Optimization Algorithm in navigating complex, high-dimensional hyperparameter search spaces and in identifying configurations that balance model expressiveness with generalization capability. The results underscore the critical role of automated hyperparameter optimization in fully exploiting the predictive potential of deep learning models for short-term electricity consumption forecasting.

From an operational standpoint, the proposed framework offers significant practical value for urban electricity distribution network operators. Accurate short-term electricity consumption forecasts can support proactive operational planning, including improved load balancing, timely congestion mitigation, and enhanced coordination of distribution assets. The ability to anticipate short-term demand variations at a fine temporal resolution is particularly valuable for reducing technical losses, preventing overload conditions, and improving overall grid reliability. Moreover, such predictive capabilities can facilitate more informed decision-making in demand-side management and contribute to the efficient integration of emerging smart grid technologies.

In a broader energy systems context, the findings of this study align with ongoing efforts toward digitalization and intelligence-driven management of power distribution networks. High-precision short-term load forecasting serves as a foundational component for advanced applications such as demand response, adaptive tariff design, and the integration of distributed energy resources. By demonstrating the effectiveness of deep learning models enhanced through metaheuristic optimization, this study provides a scalable and data-driven pathway for improving electricity demand forecasting in rapidly urbanizing regions.

Despite the strong performance achieved, several avenues for future research remain open. One important direction involves extending the proposed framework to multi-step and longer-horizon forecasting tasks, which are essential for medium-term operational planning and energy scheduling. Additionally, incorporating online or incremental learning mechanisms would allow the models to adapt continuously to evolving consumption patterns, seasonal changes, and potential structural shifts in demand behavior. Such adaptability is particularly relevant in dynamic urban environments where electricity usage patterns may change over time.

Future work may also explore the scalability and transferability of the proposed framework across different cities, distribution networks, and climatic conditions. Evaluating model performance on datasets from diverse geographic regions would provide deeper insight into the generalizability of the approach and its suitability for broader deployment. Furthermore, integrating additional explanatory variables, such as socio-economic indicators or renewable generation profiles, could further enhance forecasting accuracy and support more comprehensive energy system analysis.

Finally, the integration of the proposed forecasting framework with real-time SCADA platforms and smart grid infrastructures represents a promising direction for practical deployment. Such integration would enable near-real-time forecasting, dynamic decision support, and closed-loop control strategies for urban electricity distribution systems. By advancing toward these directions, the proposed deep learning and metaheuristic optimization framework has the potential to play a key role in the development of intelligent, resilient, and sustainable electricity networks.

Data Availability

The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/fedesoriano/electric-power-consumption.

Declarations

- **Acknowledgments**
  Not applicable.

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

References

[1] Q. Hassan et al., "The renewable energy role in the global energy transformations," *Renewable Energy Focus*, vol. 48, p. 100 545, 2024, ISSN: 1755-0084. DOI: https://doi.org/10.1016/j.ref.2024.100545. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1755008424000097.

[2] A. Sharma, S. N. Singh, M. M. Serratos, D. Sahu, and V. Strezov, "Urban energy transition in smart cities: A comprehensive review of sustainability and innovation," *Sustainable Futures*, vol. 10, p. 100 940, 2025, ISSN: 2666-1888. DOI: https://doi.org/10.1016/j.sftr.2025.100940. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666188825005052.

[3] H. Ahmad, M. Yaqub, and S. H. Lee, "Global trends in carbon neutrality: A scientometric review on energy transition challenges, practices, policies, and opportunities," *Environment, Development and Sustainability*, pp. 1–26, 2025. DOI: https://doi.org/10.1007/s10668-025-06507-7.

[4] J. Wei, Y. Wu, and M. K. Anser, "Unpacking the drivers of sustainable development: A quantile-based analysis of digitalization, resource efficiency, and economic diversification," *Environment, Development and Sustainability*, pp. 1–28, 2025. DOI: https://doi.org/10.1007/s10668-025-06751-x.

[5] A. Teshome Mossisa and A. T. Han, "Transitioning to renewable electric energy in rapidly urbanizing sub-saharan africa: Challenges and opportunities," *Energy Strategy Reviews*, vol. 61, p. 101 842, 2025, ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2025.101842. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211467X25002056.

[6] T. S. Kishore, P. U. Kumar, and V. Ippili, "Review of global sustainable solar energy policies: Significance and impact," *Innovation and Green Development*, vol. 4, no. 2, p. 100 224, 2025, ISSN: 2949-7531. DOI: https://doi.org/10.1016/j.igd.2025.100224. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949753125000219.

[7] W. Zhang and B. K. Sovacool, "The "triple burden" effect and "pressure-opportunity paradox" of net-zero transitions: Exploring the political economy of carbon border adjustment mechanism (cbam) implementation in the global south," *Energy Strategy Reviews*, vol. 62, p. 101 913, 2025, ISSN: 2211-467X. DOI: https://doi.org/10.1016/j.esr.2025.101913. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211467X25002767.

[8] S. M. RAHMAN, A. RAIHAN, M. S. ALAM, and S. CHOWDHURY, "Greenhouse gas emission dynamics and climate change mitigation efforts toward sustainability in the middle east and north africa (mena) region," *Regional Sustainability*, vol. 6, no. 4, p. 100 246, 2025, ISSN: 2666-660X. DOI: https://doi.org/10.1016/j.regsus.2025.100246. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666660X25000544.

[9] M. Imran, M. S. Alam, Z. Jijian, I. Ozturk, S. Wahab, and M. Doğan, "From resource curse to green growth: Exploring the role of energy utilization and natural resource abundance in economic development," in *Natural resources forum*, Wiley Online Library, vol. 49, 2025, pp. 2025–2047. DOI: https://doi.org/10.1111/1477-8947.12461.

[10] S. Kahil, O. Jellouli, and M. Fihri, “Hybrid clustering and extreme value theory for anomaly detection in electricity consumption: A case study on tetouan’s grid,” *Statistics*, pp. 1–42, 2025. DOI: https://doi.org/10.1080/02331888.2025.2571861.

[11] Y. Ji, Y. Zhu, S. Lu, L. Yang, and A. W.-C. Liew, “Wtc-ipst: A deep learning framework for short-term electric load forecasting with multi-scale feature extraction,” *Knowledge-Based Systems*, vol. 324, p. 113 907, 2025, ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2025.113907. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705125009530.

[12] Z. Zou, J. Wang, N. E, C. Zhang, Z. Wang, and E. Jiang, “Short-term power load forecasting: An integrated approach utilizing variational mode decomposition and tcn–bigru,” *Energies*, vol. 16, no. 18, 2023, ISSN: 1996-1073. DOI: 10.3390/en16186625. [Online]. Available: https://www.mdpi.com/1996-1073/16/18/6625.

[13] H.-M. Zuo, J. Qiu, Y.-H. Jia, Q. Wang, and F.-F. Li, “Ten-minute prediction of solar irradiance based on cloud detection and a long short-term memory (lstm) model,” *Energy Reports*, vol. 8, pp. 5146–5157, 2022, ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2022.03.182. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484722007375.

[14] W. Khan, J. Y. Liao, S. Walker, and W. Zeiler, “Impact assessment of varied data granularities from commercial buildings on exploration and learning mechanism,” *Applied Energy*, vol. 319, p. 119 281, 2022, ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2022.119281. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261922006389.

[15] E. Chaerun Nisa and Y.-D. Kuan, “Comparative assessment to predict and forecast water-cooled chiller power consumption using machine learning and deep learning algorithms,” *Sustainability*, vol. 13, no. 2, 2021, ISSN: 2071-1050. DOI: 10.3390/su13020744. [Online]. Available: https://www.mdpi.com/2071-1050/13/2/744.

[16] U. Kanewala, S. Weerakoon, and R. Nawarathna, “Exploring deep learning and tree-based ensemble models for chiller energy consumption predictions,” in *2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS)*, 2021, pp. 306–311. DOI: 10.1109/ICIIS53135.2021.9660662.

[17] S. W. Kim and Y. I. Kim, “Hybrid trrm–deep learning approach for water-cooled centrifugal chiller cop prediction,” *Case Studies in Thermal Engineering*, vol. 76, p. 107 309, 2025, ISSN: 2214-157X. DOI: https://doi.org/10.1016/j.csite.2025.107309. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214157X25015692.

[18] K. Zhu et al., “Chiller system power prediction by physical-informed neural network,” *Energies*, vol. 18, no. 23, 2025, ISSN: 1996-1073. DOI: 10.3390/en18236363. [Online]. Available: https://www.mdpi.com/1996-1073/18/23/6363.

[19] M. H. Sulaiman and Z. Mustaffa, “Chiller energy prediction in commercial building: A metaheuristic-enhanced deep learning approach,” *Energy*, vol. 297, p. 131 159, 2024, ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2024.131159. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544224009320.

[20] M. H. Sulaiman and Z. Mustaffa, “Chiller power consumption forecasting for commercial building based on hybrid convolution neural networks-long short-term memory model with barnacles mating optimizer,” *Next Energy*, vol. 8, p. 100 321, 2025, ISSN: 2949-821X. DOI: https://doi.org/10.1016/j.nxener.2025.100321. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949821X25000845.

[21] T. Sathesh and Y.-C. Shih, “Optimized deep learning-based prediction model for chiller performance prediction,” *Data & Knowledge Engineering*, vol. 144, p. 102 120, 2023, ISSN: 0169-023X. DOI: https://doi.org/10.1016/j.datak.2022.102120. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169023X22001112.

[22] K. He et al., “Predictive control optimization of chiller plants based on deep reinforcement learning,” *Journal of Building Engineering*, vol. 76, p. 107 158, 2023, ISSN: 2352-7102. DOI: https://doi.org/10.1016/j.jobe.2023.107158. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710223013384.

Received: September 14, 2025 Revised: November 25, 2025 Accepted: January 03, 2026

[23] K. He et al., "Efficient model-free control of chiller plants via cluster-based deep reinforcement learning," *Journal of Building Engineering*, vol. 82, p. 108 345, 2024, ISSN: 2352-7102. DOI: https://doi.org/10.1016/j.jobe.2023.108345. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352710223025287.

[24] H. Y. P. Napitupulu, I. G. D. Nugraha, and R. F. Sari, "Development of optimization control system for chiller plant based on a predictive model with multi stack lstm and deep learning neural network multi output," *Journal of Building Engineering*, vol. 98, p. 111 029, 2024, ISSN: 2352-7102. DOI: https://doi.org/10.1016/j.jobe.2024.111029. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235271022402597X.

[25] Z. Du et al., "Knowledge-extracted deep learning diagnosis and its cloud-based management for multiple faults of chiller," *Building and Environment*, vol. 235, p. 110 228, 2023, ISSN: 0360-1323. DOI: https://doi.org/10.1016/j.buildenv.2023.110228. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S036013232300255X.

[26] P. Li, B. Anduv, X. Zhu, X. Jin, and Z. Du, "Diagnosis for the refrigerant undercharge fault of chiller using deep belief network enhanced extreme learning machine," *Sustainable Energy Technologies and Assessments*, vol. 55, p. 102 977, 2023, ISSN: 2213-1388. DOI: https://doi.org/10.1016/j.seta.2022.102977. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138822010256.

[27] P. Li, B. Anduv, X. Zhu, X. Jin, and Z. Du, "Across working conditions fault diagnosis for chillers based on iot intelligent agent with deep learning model," *Energy and Buildings*, vol. 268, p. 112 188, 2022, ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2022.112188. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778822003590.

[28] Z. Jiang, M. J. Risbeck, S. C. Kulandai Samy, C. Zhang, S. Cyrus, and Y. M. Lee, "A timeseries supervised learning framework for fault prediction in chiller systems," *Energy and Buildings*, vol. 285, p. 112 876, 2023, ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2023.112876. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378778823001068.

[29] P. Hosseini, "Deep-learning neural network prediction of a solar-based absorption chiller cooling system performance using waste heat," *Sustainable Energy Technologies and Assessments*, vol. 53, p. 102 683, 2022, ISSN: 2213-1388. DOI: https://doi.org/10.1016/j.seta.2022.102683. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213138822007329.

[30] E.-S. M. El-Kenawy et al., "Nioa: A novel metaheuristic algorithm modeled on the stealth and precision of japanese ninjas," *Journal of Artificial Intelligence in Engineering Practice*, vol. 1, no. 2, pp. 17–35, 2024.