Hybrid Machine Learning for Electricity Load Forecasting: Integration of CatBoost and HGBoost with Optimization Techniques

Peiying Li
Mechanical and Electrical College, HanDan University, HanDan, Hebei, 056005, China
E-mail: Ylpa6_6@sohu.com

Keywords: electricity load consumption, energy demand, environmental sustainability, machine learning, cat boost, HG Boost.

Received: December 1, 2024

Electricity load consumption stands at the heart of energetic webs, being one of those main building blocks on which the very concept of power consumption in industry and regions originally relies. That includes measures and analyses of electricity consumption from single residential consumers up to large territorial levels, as well as interrelations among human activities, dependencies on technologies, and social development. It has huge implications for energy policy, planning, and environmental sustainability to gain insight into the dynamics of load consumption. Some state-of-the-art machine learning techniques, including the Cat Boost and HG Boost algorithms, are put to work in this work for forecasting electricity consumption levels. Moreover, three metaheuristic optimization algorithms—Sparrow Search Algorithm (SSA), Archimedes Optimization Algorithm (ArchOA), and Chaos Game Optimization (CGO)—were employed to fine-tune model parameters and improve forecasting outcomes. After carefully tuning hyperparameters in detail and performing broad model evaluations, six different predictive models were developed and examined. The best-performing hybrid configuration (CatBoost-CGO) achieved an R^2 of 0.999 and RMSE of 6.41 on training data, while HGBoost-SSA achieved an R^2 of 0.979 and RMSE of 37.23 on test data. These results demonstrate significant improvements over the standalone models, emphasizing the effectiveness of optimization in enhancing model convergence and prediction accuracy. The findings are discussed in detail about energy policy formulation, infrastructure planning, and sustainable development. This research is expected to further add to the contributions related to enhancing the understanding of electricity load consumption dynamics and informing better decision-making in energy.

Povzetek: Študija napovedovanja porabe električne energije uporablja CatBoost/HGBoost, okrepljena z metahevrističnim uglaševanjem, kar bistveno izboljša natančnost in ponuja uporabne vpoglede za energetsko politiko, načrtovanje infrastrukture in trajnostni razvoj.

1 Introduction

In the intricate tapestry of energy systems, electricity load consumption emerges as a cornerstone, dictating the ebb and flow of power utilization across diverse sectors and geographical regions. At its essence, electricity load consumption encapsulates the quantification and examination of the electricity consumed by various entities, be it residential households, commercial establishments, industrial facilities, or entire geographical areas, during specific time intervals. This metric serves as a barometer of energy demand, reflecting the intricate interplay between human activities, technological dependencies, and societal progress. Electricity load consumption lies at the heart of energy policy, infrastructure planning, and environmental concerns. From being the backbone of modern civilization, powering essential services, driving economic activities, and enhancing the quality of life, these rest on electricity. Detailed knowledge of electricity load consumption patterns enables decision-makers to identify infrastructure bottlenecks for targeted investment, supports energy analysts in designing demand-side management strategies such as load shifting and peak shaving, and allows stakeholders to allocate energy resources more efficiently based on usage trends. Understanding the dynamics of load consumption is not only essential for operational planning but also plays a critical role in addressing climate change and integrating renewable energy sources, where variability in both supply and demand must be managed to ensure energy resilience and support the transition toward low-carbon systems. The drivers for electricity consumption are multifaceted, emanating from socio-economic dynamics, climatic variations, technological changes, and regulatory frameworks. Seasonal variations, weather patterns, and geographical characteristics predominantly influence the consumption pattern. Peak loads normally fall during those periods of the year when extreme weather conditions or seasonally high energy consumption is usually recorded. The socio-economic factors include population density, rates of urbanization, and industrial activities that influence the quantum and dispersion of electricity consumption in communities. Furthermore, policy interventions, energy efficiency measures, and technological innovations greatly contribute to molding consumption patterns toward sustainable energy transitions. Understanding the complexity of electricity load consumption has great implications for formulating energy policies, infrastructure planning, and sustainable development. This load consumption data will help policymakers provide effective demand management programs, grid optimization, and incentives for energy conservation practices. It also informs land use decisions by urban planners, encourages energy-efficient building design, and helps to create more resilient communities. From a broader perspective, integrating load consumption considerations into long-term energy planning frameworks facilitates the alignment of energy investments with climate targets, fosters innovation in energy technologies, and enhances energy security at national and global scales. Electricity serves as the cornerstone in sustaining the highly technologically advanced industrialization witnessed across global economies [1][2][3]. Virtually every facet of modern activity is contingent upon the availability of electricity. As the years progress, the demand for and utilization of electric energy continues to escalate on a global scale. However, the processes involved in generating, transmitting, and distributing electrical energy remain intricate and costly. Consequently, effective grid management emerges as a pivotal function in mitigating the expenses associated with energy production while concurrently bolstering generating capacity to align with the burgeoning demand for electric energy [4]. In essence, effective grid management encompasses a multifaceted approach involving meticulous load demand forecasting, comprehensive maintenance scheduling for generating, transmission, and distribution infrastructure, and the seamless distribution of loads through supply lines. Thus, the implementation of accurate load forecasting methodologies assumes paramount importance in optimizing the efficiency of planning processes within the power generation sector [5]. Large buildings constitute a significant portion of global electricity consumption, accounting for approximately 32% of total energy usage worldwide. In Europe, these structures contribute to 36% of CO2 emissions [6]. Meeting the energy efficiency goals set for Europe by 2020, which include a 20% improvement in energy efficiency and a 20% reduction in greenhouse gas emissions from 1990 levels [7], underscores the growing importance of Building Management Systems (BMS). These systems have garnered increasing interest due to their potential to achieve energy savings and minimize environmental impact while meeting operational objectives at minimum energy costs [8][9]. Of all the BMS functionalities, load forecasting represents a very important functionality and a frequent challenge in electrical distribution systems. Various works have been reported on the study of short-term load forecasting by several techniques, which include statistical models, regression methods, state-space methods, evolutionary programming, fuzzy systems, and artificial neural networks (ANN) [10][11][12][13]. ANN is the most well-known among them due to its interpretable model, ease of implementation, and, above all, better performance. Load forecasting is a basic input to energy studies in utility systems and is especially useful when medium- and low-voltage energy distribution systems are to be studied. This concept not only helps in operational efficiency but in understanding the economics of the system to some extent— ahead-of-time opportunity and risk analysis. The evolving energy market, along with the adoption of smart grid paradigms, emphasizes advanced load management strategies and more reliable forecasts, from the single end-user up to the system scale. In addition, coordinated load forecasting becomes even more crucial since integrating uncertain renewable energy generation with actual demand is of increasing importance [14]. It is a scientific discipline within the broad area of Artificial Intelligence; its objective is to build systems capable of learning autonomously. In the case of this field, learning means to identify complex patterns in large volumes of data. The core of the learning here is the use of algorithms that analyze data in the prediction of future behaviors. Here, the term "automatically" means that, in living systems, such a process improves iteratively with time and without human intervention [15][16]. The advantage of machine learning allows the analysis of large datasets that are either too difficult or impossible for a human to go through, hence obtaining interesting insights from them. These algorithms, allowed through inherent variables in datasets, bring forth patterns of behaviors that are identified and then use those to explain which factors are most responsible for changes in these behavioral trends [17][18][19]. In energy, Machine Learning allows energy trading companies to forecast fluctuations in the consumption of electricity by their customers and thus offer them appropriate tariffs or efficiently manage energy provisioning. That means that Machine Learning enables one to shift from reactive to proactive approaches as far as energy management is concerned. By deploying well-structured and organized historical client data, an efficient database is created that could even allow one to foresee future behaviors in energy consumption. Therefore, the firms can adjust their marketing strategies according to new tariffs or optimize supply chains based on high demands for energy [20][21][16][22]. Power system management development has drawn an increasingly important academic interest in the forecast of electricity consumption. An accurate forecasting model may lead users to adjust their consumption patterns to market fluctuations and move toward a homogeneous rational expectation equilibrium [23][24]. This, in turn, ensures an uninterrupted power supply to consumers [25]. Besides, the power industry can function in an environmentally friendly manner with the assistance of appropriate electricity consumption forecasting models that will help them modify the production and consumption pattern of electricity. It will also help in the decision-making process about energy management policies regarding load unit commitment, operational security of electrical plants and power systems, economic load dispatching, and energy marketing [26][27]. Such models not only bridge the gap between electricity supply and demand but also significantly enhance the operational efficiency of power grid management systems [28][29]. Furthermore, accurate forecasting facilitates the establishment of power system schedules, thereby bolstering operational security [30]. The accurate forecast will enable customers who benefit from the forecast to plan their power usage so that the usage can be planned to balance the peak and off-peak periods, thus indirectly contributing to the reliability of the system [31]. Notably, research suggests that even a 1% improvement in load forecasting accuracy in both China and the UK yields substantial financial gains [32]. However, electricity consumption data present a challenge due to their nature as a random, non-stationary sequence influenced by various factors, including economic indicators. This poses significant hurdles to accurate forecasting [33]. Especially Gulf Cooperation Council (GCC) countries have experienced fast economic growth in the last couple of decades with a growing electricity demand accordingly. AlKhars 2019 explains how different problems related to the demand and supply of electricity in GCC countries are modeled using seven different quantitative analytical techniques. Given this background, the following paper seeks to delve deep into the most widely used Machine Learning models for forecasting electricity consumption. Apart from this, this review intends to provide comprehensive knowledge of key variables that affect the pattern of energy consumption. The primary objective of this study is to improve the precision, convergence, and generalizability of electricity load forecasting models by employing a hybrid framework that combines gradient boosting techniques with advanced metaheuristic optimization algorithms. This integration is designed to address the limitations of standalone models in capturing nonlinear consumption patterns and adapting to complex data structures. By systematically evaluating different hybrid configurations, the research aims to demonstrate the potential of such models in delivering more reliable forecasts, which are crucial for informed energy planning and policy formulation.

In recent years, a growing number of studies have explored machine learning-based approaches to electricity load forecasting, highlighting various algorithms, data sources, and optimization techniques. Despite this, challenges related to predictive accuracy, adaptability to nonlinear patterns, and model convergence remain open. To justify the need for the current study, a comparison of relevant state-of-the-art models is presented in Table 1. This comparative overview underscores the novelty of integrating CatBoost and HGBoost with metaheuristic optimization methods, as introduced in this paper.

Table 1: Comparative overview of state of the electricity load forecasting models

Model	Dataset	Evaluation Metrics	Main Findings
Ahmad & Chen [23]	Utility company data	MAE, RMSE	Compared multiple ML models; Random Forest showed reliable short-term forecasting performance.
Ahmad & Chen [24]	Utility load data	RMSE, MAPE	Nonlinear Autoregressive and Random Forest methods applied; RF performed better in non-linear contexts.
Raza et al. [28]	Smart grid load data	RMSE, MAE	Developed a hybrid ANN-ARIMA model; outperformed individual models for short-term predictions.
Xiao et al. [29]	Seasonal load data	RMSE	Introduced a combined model using seasonal patterns and a modified firefly algorithm; improved accuracy.
Massana et al. [30]	Smart City platform	MAPE	Used data-driven models; highlighted importance of identifying service-specific patterns in forecasting.
Hong [32]	Various datasets	Various	Offered a comprehensive overview of hybrid intelligent systems for energy demand forecasting.

1.1 Main contribution

It proposes advanced machine learning methodologies that could predict the consumption of electricity load (KW), with a major emphasis on enhancing the accuracy of the prediction. The research integrated advanced techniques like CatBoost and HGBoost, as well as the inclusion of different nature-inspired optimization algorithms, such as SSA, ArchOA, and CGO. This approach also conducted an extended comparative analysis regarding the performance of machine learning methods in individual and combined ways. This study represents a significant contribution because it attempts to contribute to the literature regarding the application of machine learning for efficient forecasts of the consumption of electricity load in KW. This helps an individual entity draw very useful inferences from the cumulative gains of forecast precision through an individual as well as an integrated model approach. The subsequent sections of this research paper follow the proper academic track. Section 2 provides a thorough discussion of the techniques used for prediction. This includes an overall detail of the proposed models, CatBoost and HGBoost, together with a description of how the optimization techniques work. Results and further analysis of the models are presented in Section 3. For the sake of deep examination, this section is complemented with several charts and tables. Finally, Section 4 concludes the paper by summarizing the principal findings and implications derived from this study. The systematic approach followed in this research contributes to the academic discourse on applying machine learning in the forecasting of electricity consumption, setting the base for future research attempts.

2 Methodology

The main emphasis of this paper is on using the advanced machine learning techniques of the CatBoost and HGBoost algorithms to predict the level of electricity consumption. Another important aspect of this investigation is improving the predictive capabilities of these models by very careful hyperparameter optimization. For this purpose, some sophisticated optimization algorithms known as SSA, ArchOA, and CGO will be applied to optimize the key parameters governing the performance of the best-performing model. The current research will focus on the development and validation of six different predictive models for forecasting levels of electricity consumption. The result of an extensive comparative analysis of the models through robust statistical evaluation techniques will, therefore, help in the identification of the most efficient and reliable predictive models out of the considered variants. The methodological framework involves extensive data collection, comprehensive analysis, and stringent data validation processes. Immediately after data validation, an informed partitioning scheme is introduced, where the dataset is split into 80% training and 20% testing. The partitioning scheme is important to be used in ensuring that the predictive models are robust and accurate since this allows for a comprehensive test of their performances on different datasets. This may be a two-stage process: CatBoost and HGBoost algorithms were separately tested to make independent predictions, and then the research focused on hybrid models encapsulating the strengths of CatBoost and HGBoost for better predictive accuracy. Its refinement in terms of accuracy and efficiency has been sought through a painstaking process of optimization of hyperparameters systematically and intensively. It is in this regard that the application of various optimization techniques, including SSA, ArchOA, and CGO, has been quite instrumental in selecting the best hyperparameter settings for both CatBoost and HGBoost models. A similar optimization strategy is very instrumental in enhancing the predictive performance and reliability of forecasts provided by the models in forecasting electricity consumption. The dissemination of the study's findings will be facilitated through comprehensive discussions in subsequent sections, complemented by insightful graphical representations, charts, and tables. These visual aids are strategically integrated to provide a clear and comprehensive understanding of the research outcomes, promoting deeper comprehension and interpretation within the academic community. Additionally, Fig 1 presents a schematic illustration outlining the comprehensive structure of the employed model along with the methodology applied in this research, depicted as a flowchart. This scholarly inquiry primarily focuses on investigating the geographical area of a campus-distributed energy system in Japan, spanning a temporal range from 2020 to 2021.

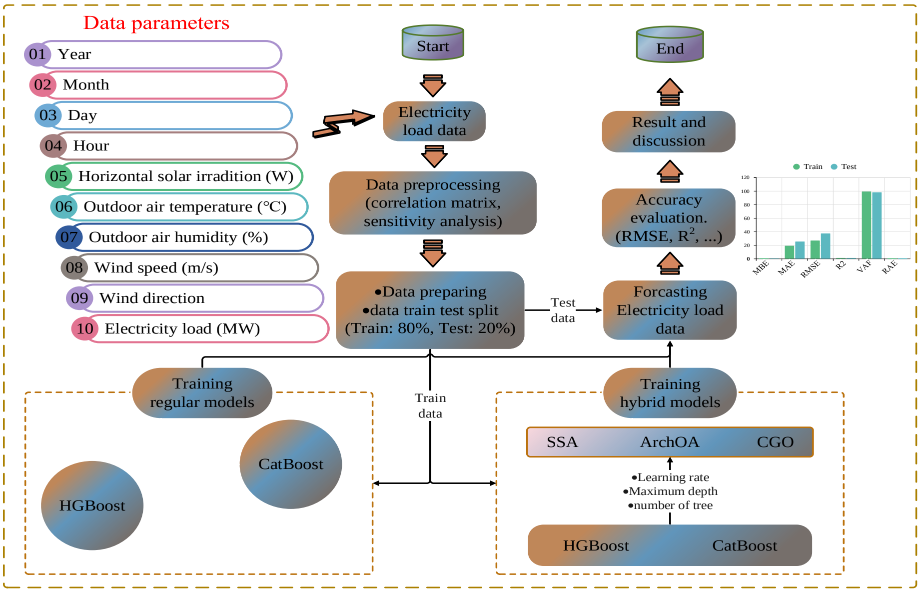

Figure 1: The flowchart diagram of the current investigation

2.1 Data

The dataset employed in this study originates from previous investigations conducted by Beeravalli and JIN, Xiaoyu [34]. This scholarly inquiry primarily focuses on examining the geographical reach of a campus-based distributed energy system in Japan, spanning the time frame from 2020 to 2021. It is imperative to acknowledge that the estimation of electricity usage is subject to numerous significant factors. To understand the underlying dynamics and justify the selection of input variables for our predictive models, a preliminary statistical analysis was performed. This process began with an Exploratory Data Analysis (EDA) to investigate the relationships between the potential predictor variables and the target variable, electricity load (kW). A correlation analysis was conducted to quantify the strength and direction of these relationships. Key findings from this analysis indicated strong correlations between electricity load and variables such as outdoor air temperature and horizontal solar irradiation, which aligns with expected physical principles of energy consumption for cooling and lighting. Other variables like humidity and wind speed also showed relevant, albeit weaker, correlations.

This examination confirmed the suitability of these variables for forecasting electricity consumption. Table 2 provides the descriptive statistics for the final set of input variables used in this study, summarizing their central tendency, dispersion, and range. It establishes an empirical foundation for the subsequent analyses and findings presented in this research. By systematically delineating influential factors within a structured table, this study sets the groundwork for a thorough comprehension of the intricate interplay among various elements affecting the predictive modeling of electricity consumption.

Table 2: The input parameters and their corresponding statistical particulars

variables	count	mean	std	min	25%	50%	75%	max
Year	17544	2020.499	0.500014	2020	2020	2020	2021	2021
Day	17544	15.73871	8.804172	1	8	16	23	31
Month	17544	6.519836	3.449649	1	4	7	10	12
Hour	17544	11.5	6.922384	0	5.75	11.5	17.25	23
Horizontal solar irradiation (W)	17544	165.0691	263.2195	0	0	2	238	1280
Outdoor air temperature (℃)	17544	17.18443	7.798283	-2.3	10.8	17.1	24	36.5
Outdoor air humidity (%)	17544	68.63321	15.00121	17	57	69	80	100
Wind speed (m/s)	17544	2.164774	1.736671	0	1	1.7	2.9	14.8
Wind direction	17544	265.0778	120.6902	0	170	243	374	535
Electricity load (kW)	17544	663.9937	255.0504	218	471	563	828	1553
Year	17544	2020.499	0.500014	2020	2020	2020	2021	2021

2.2 Machine learning methods

An academic overview of methodologies applied for forecasting the consumption of electricity is carried out in this section. To serve this purpose, advanced models at different steps are strategically utilized, which include the CatBoost and HGBoost models. Furthermore, optimization techniques such as SSA, ArchOA, and CGO are incorporated into the research framework for increasing efficiency and building better predictive accuracy. The selection of such methodologies should be done with due care, as this deeply influences the strength and accuracy of predictions related to electricity consumption. This important point will be elaborated upon in the following sections of the given comprehensive study. This methodological approach underlines the sophistication not only of the models used but also of the conscious effort towards performance optimization to make the entirety of the reliability and effectiveness of the process of forecasting electricity consumption rich.

2.2.1 Categorical gradient boosting (Catboost)

CatBoost, a notable advancement in gradient boosting and decision tree frameworks, holds significant relevance within academic discourse [35],[36]. Boosting generally works according to the principle of combining weak models into one robust predictive model that is only a little bit better than random guessing. In the case of gradient boosting, the systematic reduction of errors is performed by building decision trees sequentially, each learning from the errors of the last-but-one tree until the selected loss function has reached a minimum value. What distinguishes it from other, more traditional gradient-boosting models is a unique form of decision tree construction. Whereas the trees in traditional gradient-boosting involve ordered comparisons of predicted values, CatBoost makes use of "oblivious trees." In these trees, all nodes on the same level compare the same predictor for the same value. The CatBoost algorithm, further, follows a different approach of choosing the data to fit the ht+1 decision tree by using an arbitrary ordering and randomly permuting the components of D. This process involves the creation of a new set Dk = {x1, x2..., xk-1}, where x1, x2..., xk-1 represent the elements of D sorted according to the random permutation σ, and (k) denotes the kth element of D within the permutation σ's context. In contrast to strictly adhering to Eq. (1), CatBoost utilizes a modified version in its analysis to determine the encoded value Xk for the ith categorical value during the fitting of Decision Tree ht+1.

$$\hat{X}_k^i = \frac{\sum x_j \in D_k 1x_j^i = x_k^i . y_j + ap}{\sum x_j \in D_k 1x_j^i = x_k^i + a} \tag{1}$$

2.2.2 Histogram gradient boosting regressor (HGBoost)

HGBoost, also known as Histogram Gradient Boosting Regressor [37], presents a modified approach to gradient boosting regression by incorporating histograms for splitting, departing from conventional methods. By using histograms, HGBoost efficiently quantizes the feature space into many bins, hence allowing much finer and more flexible partitions. That also makes the HGBoosting algorithm very useful when coping with categorical and continuous variables since this increases the granularity with which these variables are partitioned. One of the most appealing aspects of HGBoost is its lower sensitivity to the choice of hyperparameters when compared to the standard gradient-boosting regressor. This makes it less sensitive, adding to its strength and reducing the tendency of overfitting while training the model. Another point that could be mentioned here is that the use of histograms increases the computational cost related to the training process in HGBoost. Overall, it makes HGBoost an enhanced, flexible approach toward regression problems using histograms for segmentation and giving better approximations. Its ability to support a wide range of variable types and its lower dependence on hyperparameters make it very useful for many purposes in machine learning.

2.2.3 Sparrow search algorithm (SSA)

The Sparrow Search Algorithm (SSA) was initially proposed by Xue Jiankai [38], drawing inspiration from the foraging and anti-predation behaviors observed in sparrow populations. Sparrows, being a diverse and widespread avian group, inhabit various geographical regions globally. They exhibit a notable ability to adapt to human settlements, often coexisting in areas populated by humans. Their diet primarily comprises seeds derived from grains or herbs. Sparrows demonstrate remarkable cognitive abilities and possess a robust memory. Within the sparrow population, two distinct behavioral types are observed: producers and scroungers. Producers actively engage in locating food sources, while scroungers depend on the efforts of producers for sustenance. Producers demonstrate cautious behavior and identify foraging areas or directions beneficial to all scroungers, playing a pivotal role in identifying regions abundant in food sources. Although a sparrow can transition into a producer by actively seeking optimal food sources, the ratio of producers to scroungers remains constant within the population. Scroungers rely on producers to discern and provide the most suitable food sources. The position of the producer is updated through the following procedure:

$$X_{i,j}^{t+1} = X_{i,j}^{t}.exp(-\frac{1}{\alpha.iter_{max_2}} \qquad X_{i,j}^{t+1} = X_{i,j}^{t} + Q.LifR_2 \geq ST \tag{2}$$

In this formulation, the updated value of the jth dimension of the ith sparrow in the subsequent iteration is denoted as $X_{i.j}^{t+1}$. The maximum number of iterations is designated as itermax. The parameter α is a uniformly distributed random variable ranging from 0 to 1. The variables R2 (lying within the range of 0 to 1) and ST (ranging between 0.5 and 1.0) correspond to the values associated with fap (false alarm probability) and the threshold for solidity, respectively. Additionally, Q represents another random variable employed to ensure a normal distribution. L is a one-dimensional matrix of size 1×d, with each element initialized to 1. In scenarios where the value of R2 falls below ST, indicating the absence of predators in the vicinity, the sparrow adopts a strategy of broad search. Conversely, if R2 exceeds or equals ST, signifying the presence of predators detected by certain sparrows, all sparrows are mandated to relocate rapidly (with a velocity denoted by fy) to safer zones [39].

2.2.4 Archimedes optimization algorithm (ArchOA)

The Archimedes Optimization Algorithm (AOA) is a meta-heuristic algorithm inspired by principles from physics, specifically focusing on the buoyant force experienced by objects submerged in fluids, which is equal to the weight of the displaced fluid [40]. This discussion aims to provide a comprehensive understanding of the mathematical foundations that underlie the AOA.

$$O_i = lb_i + rand\,^*(ub_i - lb_i), \forall i \tag{3}$$

$$den_i\,, vol_i = rand, \forall i \tag{4}$$

$$acc_i = lb_i + rand^*(ub_i\,\_lb_i), \forall i \tag{5}$$

In the given mathematical expressions, let O_i be the i-th object of the population, and let lb and ub be the lower and upper bounds of the search space. The parameters den, vol, acc, and rand correspond respectively to properties associated with density, volume, acceleration, and a randomly generated number within the interval [0, 1]. These equations play a pivotal role in determining the initial positions (x_{best}) for the entire population. Specifically, x_{best}, den_{best}, vol_{best}, and acc_{best} are computed based on the evaluation of each position's performance.

$$den_i^{t+1} = den_i^t + rand^*(den_{best} - den_i^t), \forall i \tag{6}$$

$$vol_i^{t+1} = vol_i^t + rand^*\left(vol_{best} - vol_i^t\right), \forall i \tag{7}$$

$$TF = exp\left(\frac{t - t_{max}}{t_{max}}\right) \tag{8}$$

$$d^{t+1} = exp\left(\frac{t_{max} - t}{t_{max}}\right) - \left(\frac{t}{t_{max}}\right) \tag{9}$$

$$ifTF \leq 0.5 \to acc_i^{t+1} = \frac{den_k + vol_k * acc_k}{den_i^{t+1} * vol_i^{t+1}} \tag{10}$$

$$ifTF > 0.5 \to acc_i^{t+1} = \frac{den_{best} + vol_{best} * acc_{best}}{den_i^{t+1} * vol_i^{t+1}} \tag{11}$$

$$acc_{i-norm}^{t+1} = 0.9 * \frac{ass_i^{t+1} - \min(acc)}{\max(acc) - \min(acc)} + 0.1 \tag{12}$$

$$ifTF \leq 0.5 \to x_i^{t+1} = x_i^{t+1} + c_1 * rand * acc_{i-norm}^{t+1} * d * (x_{rand} - x_i^t) \tag{13}$$

$$ifTf > 0.5 \to x_{best}^t + F * c_2 * rand * acc_{i-norm}^{t+1} * d * (c_3 * TF * x_{best} - x_i^t) \tag{14}$$

$$F = \begin{cases} +1if2 * rand - c_4 \leq 0.5 \\ -1if2 * rand - c_4 > 0.5 \end{cases} \tag{15}$$

In this context, the variables den_{best}, vol_{best}, and $rand$ correspond to the density and volume of the bestdiscovered object thus far, as well as a uniformly distributed random number, respectively. The parameters TF, d, t, and t_{max} represent the transfer operator, density factor, the current iteration, and the maximum iteration count, respectively. Additionally, den_k, vol_k, and acc_k signify the density, volume, and acceleration of the kth object chosen randomly from the population, while acc_{i-norm}^{t+1} denotes the normalization of acc_i^{t+1}, and x_i^{t+1} represents the subsequent position of object i. Constants C_1, C_2, C_3, and C_4 are coefficients, and the flag F is utilized to alter the direction of motion. The equilibrium between the exploration and exploitation phases is regulated by the transfer operator (TF) and the density factor (d). Here, the concept of the transfer operator can be elaborated with the example of two objects submerged in a fluid when objects are initially hitting each other and finally reach equilibrium. In this case, the exploration phase depends on the factor that if the transfer operator is less than 0.5, then equations 15 and 18 imply the collision between two agents of a population. On the other hand, the objects reach an equilibrium state where the start of the exploitation phase utilizes equations 11 and 14.

2.2.5 Chaos game optimization (CGO)

The study of chaotic systems has been a domain of great scholarly interest, mainly because of the very interdisciplinary nature the concept of chaos carries within itself [41]. A very interesting application in this realm is the inclusion of chaotic systems within optimization algorithms, as reflected by the introduction of a new metaheuristic called Chaos Game Optimization (CGO). Fundamentally, the essence of the CGO algorithm encompasses a basic principle based on chaos theory, wherein fractal generation is enabled by exploiting the chaos game technique. The very basic underlying principle of the methodology of CGO deals with the theory behind chaos games that consecutively shapes the operational feature set of the algorithm. The CGO algorithm introduces a set of candidate solutions (X) that act like seeds. Each individual solution candidate x_i is made up of configurable variables x_ (i, j) that represent the position of the seed. The algorithm searches the Sierpinski triangle for solution space. The subsequent section presents the mathematical representation of this exploration.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^j & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^j & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_i^1 & x_i^2 & \cdots & x_i^j & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^j & \cdots & x_n^d \end{bmatrix}, \begin{cases} i = 1,2,\ldots,n. \\ i = 1,2,\ldots,d. \end{cases} \tag{16}$$

In this context, 'n' represents the magnitude of the population within the exploration field, whereas 'd' refers to the dimensionality of the problem. The seeds are initially allocated randomly using the subsequent procedure:

$$x_i^j(0) = x_{i,\min}^j + \text{rand.}\left(x_{i,\max}^j - x_{i,\min}^j\right), \begin{cases} i = 1,2,\ldots,n. \\ i = 1,2,\ldots,d. \end{cases} \tag{17}$$

In this context, $x_i^j(0)$ signifies the initial population, with $x_{i,\min}^j$ and $x_{i,\max}^j$ representing the minimum and maximum values for the jth design variable, respectively. The term 'rand' denotes a randomly generated number falling within the range of 0 to 1. To define the configuration of a triangle, the mathematical model produces various starting points within their designated upper and lower limits.

Figure 2 presents the pseudo-code representations of the three metaheuristic optimization algorithms utilized in this study: Chaos Game Optimization (CGO), Sparrow Search Algorithm (SSA), and Archimedes Optimization Algorithm (ArchOA). These algorithms were selected due to their proven efficiency in handling complex search spaces and their ability to enhance the convergence and accuracy of machine learning models through hyperparameter tuning. Each algorithm incorporates unique mechanisms—chaotic dynamics in CGO, social foraging behavior in SSA, and physical principles of buoyancy and displacement in ArchOA—that contribute to effective exploration and exploitation of the solution space.

Algorithm 2 Archimedes Optimization Algorithm (ArchOA)

Input: Population size N, max iterations $T_{\max}$, bounds $[lb, ub]$
Output: Optimal solution X_{best}
Initialize objects $O_i = lb_i + rand \cdot (ub_i - lb_i)$
Initialize $den_i, vol_i \sim U(0,1)$
Initialize $acc_i = lb_i + rand \cdot (ub_i - lb_i)$
Evaluate fitness, set O_{best}
for $t = 1$ to $T_{\max}$ **do**
  $TF = \exp\left(\frac{t - T_{\max}}{T_{\max}}\right)$
  $d = \exp\left(\frac{T_{\max} - t}{T_{\max}}\right) - \frac{t}{T_{\max}}$
  for each object i **do**
    Update $den_i^{t+1} = den_i^t + rand \cdot (den_{\text{best}} - den_i^t)$
    Update $vol_i^{t+1} = vol_i^t + rand \cdot (vol_{\text{best}} - vol_i^t)$
    if $TF \leq 0.5$ **then**
      Randomly select $k \neq i$
      $acc_i^{t+1} = \frac{den_k + vol_k \cdot acc_k}{den_i^{t+1} \cdot vol_i^{t+1}}$
      $X_i^{t+1} = X_i^t + C_1 \cdot rand \cdot acc_i^{t+1} \cdot d \cdot (X_{\text{rand}} - X_i^t)$
    else
      $acc_i^{t+1} = \frac{den_{\text{best}} + vol_{\text{best}} \cdot acc_{\text{best}}}{den_i^{t+1} \cdot vol_i^{t+1}}$
      $X_i^{t+1} = X_{\text{best}}^t + F \cdot C_2 \cdot rand \cdot acc_i^{t+1} \cdot d \cdot (T \cdot X_{\text{best}} - X_i^t)$
      $(F = \begin{cases} +1 & \text{if } rand > 0.5 \\ -1 & \text{otherwise} \end{cases})$
    end if
  end for
  Apply boundary constraints
  Update O_{best}
end for

(a)

Algorithm 1 Sparrow Search Algorithm (SSA)

Input: Population size N, max iterations $T_{\max}$, alarm threshold $ST \in [0.5, 1]$, producer ratio PD
Output: Optimal solution X_{best}
Initialize sparrow positions X_i $(i = 1, 2, \ldots, N)$
Classify sparrows: producers $(N_{prod} = PD \times N)$, scroungers $(N - N_{prod})$
for $t = 1$ to $T_{\max}$ **do**
  for each producer i in $[1, N_{prod}]$ **do**
    Generate random $R_2 \in [0,1]$, $Q \sim \mathcal{N}(0,1)$
    Update position:
$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(-\frac{t}{\alpha \cdot T_{\max}}\right) & \text{if } R_2 < ST \\ X_{i,j}^t + Q \cdot L & \text{if } R_2 \geq ST \end{cases}$$
    (where $\alpha \in (0,1]$, $L =$ unit vector)
  end for
  for each scrounger i in $[N_{prod} + 1, N]$ **do**
    Update position:
$$X_{i,j}^{t+1} = Q \cdot \exp\left(\frac{X_{\text{worst}}^t - X_{i,j}^t}{i^2}\right)$$
  end for
  for each boundary-aware sparrow (random 10-20% of population) **do**
    Adjust position:
$$X_{i,j}^{t+1} = X_{\text{best}}^t + \beta \cdot |X_{i,j}^t - X_{\text{best}}^t|$$
    $(\beta \sim \mathcal{N}(0,1),\ K \in [-1,1])$
  end for
  Evaluate fitness $f(X_i^{t+1})$
  Update X_{best}, X_{worst}
end for

(b)

Algorithm 3 Chaos Game Optimization (CGO)

Input: Population size N, max iterations $T_{\max}$, chaos parameter $\beta = 2.5$
Output: Optimal solution X_{best}
Initialize seeds $x_i = lb_i + rand \cdot (ub_i - lb_i)$
Evaluate fitness, set X_{best}
for $t = 1$ to $T_{\max}$ **do**
  for each seed i **do**
    $GP_1 = X_{\text{best}}$
    $GP_2 = \text{mean}(X)$
    $GP_3 = X_i$
    Generate $r \sim U(0,1)$
    if $r < 0.33$ **then**
      $X_{\text{new}} = GP_1 + \beta \cdot (GP_1 - GP_3)$
    else if $r < 0.66$ **then**
      $X_{\text{new}} = GP_2 + \beta \cdot (GP_2 - GP_3)$
    else
      $X_{\text{new}} = GP_3 + \beta \cdot (GP_3 - \text{mean}(X))$
    end if
    Evaluate $f(X_{\text{new}})$
    if $f(X_{\text{new}}) < f(X_i)$ **then**
      $X_i^{t+1} = X_{\text{new}}$
    end if
  end for
  Update X_{best}
end for

(c)

Figure 2: Illustration of the pseudo-codes of the employed metaheuristic optimization algorithms: (a) Chaos Game Optimization (CGO), (b) Sparrow Search Algorithm (SSA), and (c) Archimedes Optimization Algorithm (ArchOA)

2.3 Model verification and evaluation

To validate the proposed models, a range of performance metrics and analytical methods is employed. These metrics aim to detect disparities between observed and predicted values by evaluating residual errors. The utilized metrics include (MBE), (MAE), (RMSE), (R2), (VAF), and (RAE) [42],[43]. The precise mathematical expressions for these statistical measures are outlined in Table 3.

Table 3: Statistical evaluation indexes

Statistics	Criteria	Equation
VAF	Variance Accounted For	$100\% \times \frac{\sum_{i=1}^{n}(y_i - \bar{y})(f_i - \bar{f})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
RMSE	Root Mean Square Error	$\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$
MBE	Mean Bias Error	$\frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)$
MAE	Mean Absolute Error	$\frac{\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert}{n}$

R2	Coefficient of Determination	$1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$
RAE	Relative Absolute Error	$\frac{[\sum_{i=1}^{n}(\hat{y}_i-y_i)^2]^{\frac{1}{2}}}{[\sum_{i=1}^{n}(y_i)^2]^{\frac{1}{2}}}$

3 Results and discussion

In this section of the academic manuscript, we explore the results obtained by employing both individual and combined modeling approaches to predict electricity consumption. The individual models under examination include CatBoost and HGBoost techniques. In contrast, the combined models integrate these individual algorithms, bolstered by a fusion of three distinct optimization strategies: SSA, ArchOA, and CGO. The empirical outcomes from these experiments are presented systematically using various charts, visual representations, and meticulously organized tables. Other sections of this paper will critically analyze these results, along with a comprehensive discussion and analytical assessment, to be in a position to fully cover the results obtained from the research.

Fig 3: Complete construction of a correlation matrix, including input and output variables taken into our model. Moreover, the input parameters that will be used are Year, Day Month, Hour, Horizontal solar irradiation (W), Outdoor air temperature (C), Outdoor air humidity (%), Wind speed (m/s), and Wind direction, while the output to be estimated in this analysis will be gas load (m3). The color gradient within the chart shows from -0.36 to +1. First of all, it is important to point out that the negative sign means the inverse correlation, while the positive sign means direct. Having analyzed the picture, it could be noticed that some parameters express a negative correlation with the dependent variable, like Day and Outdoor air humidity. On the other hand, Horizontal solar irradiation (W) and Outdoor air temperature (C) are positively related to the target variable. That would mean if these measures are increased, so is the output energy, which can be interpreted from the visual representation.

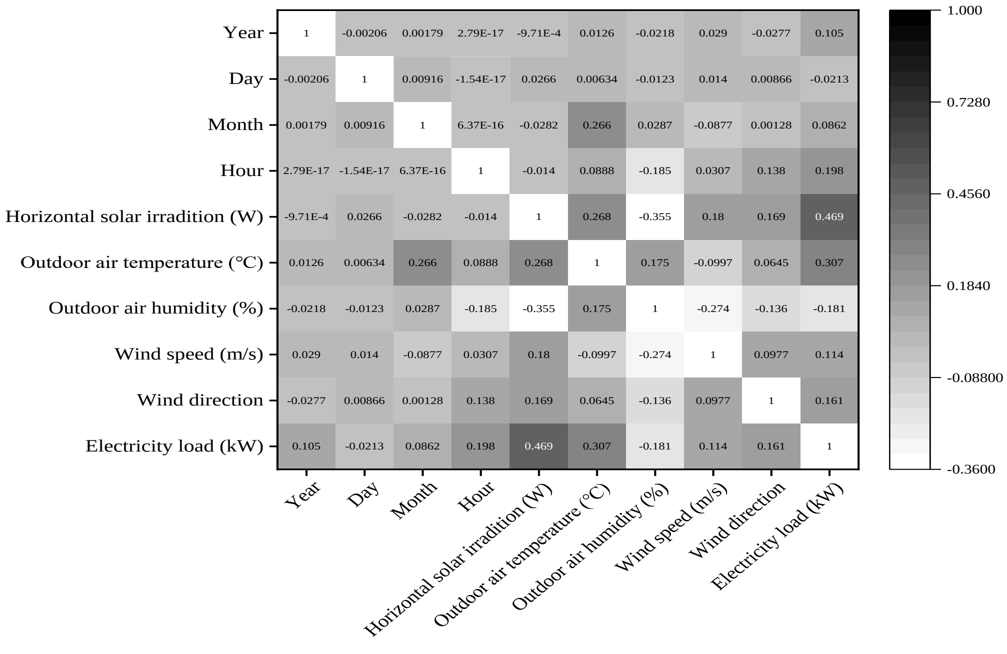

Figure 3: The correlation matrix of features

Fig 4: Sensitivity and input importance plot regarding the model predictions over electricity consumption. Such graphical visualizations help to make a more insightful analysis of the relative contribution of various model parameters on the overall results of predictions. In particular, parameters such as "Hour", "Horizontal solar irradiation (W)", and "Outdoor air temperature (C)" turn out to be more contributive with higher sensitivity and a larger impact on model predictions. By contrast, the input parameters "Day," "Wind direction," "Wind speed," and "Outdoor air humidity" are less sensitive; that would reflect their relatively little impact on the model's outputs when compared against the model's other input variables.

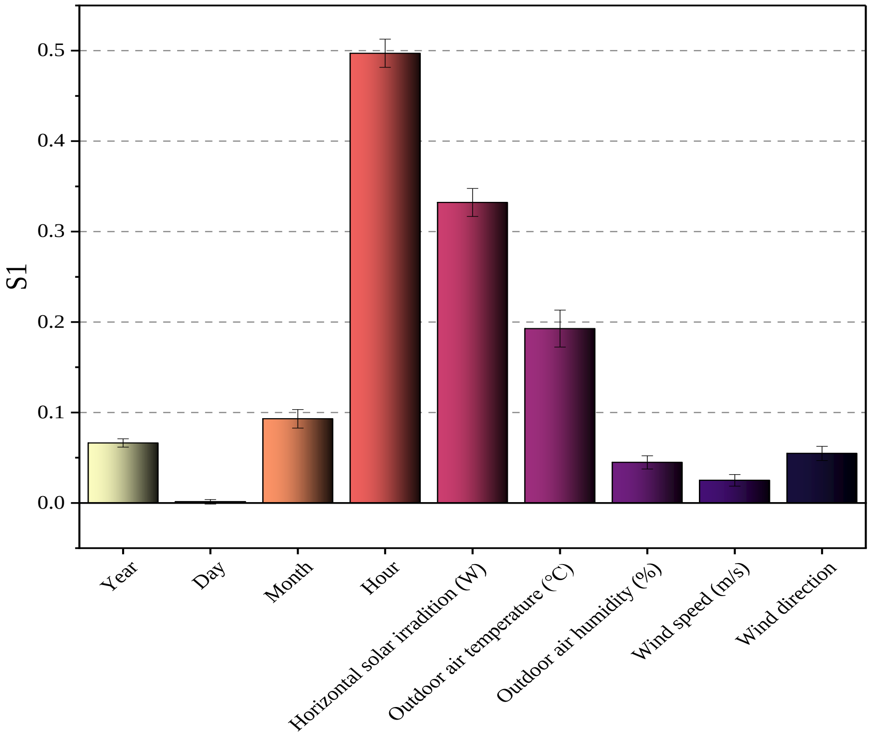

Figure 4: Sensitivity analysis of variables

Fig 5 summarizes detailed information about the results from each model: a time series trend of data, scatterplot, and performance evaluation metrics for the training and testing phases. The temporal illustration in the time series graph vividly portrays the error rate in purple, emphasizing the commendable performance of both models during the training phase. However, upon evaluation against the test dataset, the CatBoost model exhibits comparatively superior performance when juxtaposed with the alternative model. A meticulous analysis of the scatterplot and pertinent statistical indicators, particularly focusing on the coefficient of determination (R2) depicted in the graphical representation, elucidates that the CatBoost algorithm demonstrates heightened performance, attaining an R2 value of 0.6823 in contrast to its counterparts. These findings underscore the predictive precision and resilience of the CatBoost algorithm in the domains of modeling and forecasting within an academic context. For a better comparison of the models, the results are illustrated in Table 4.

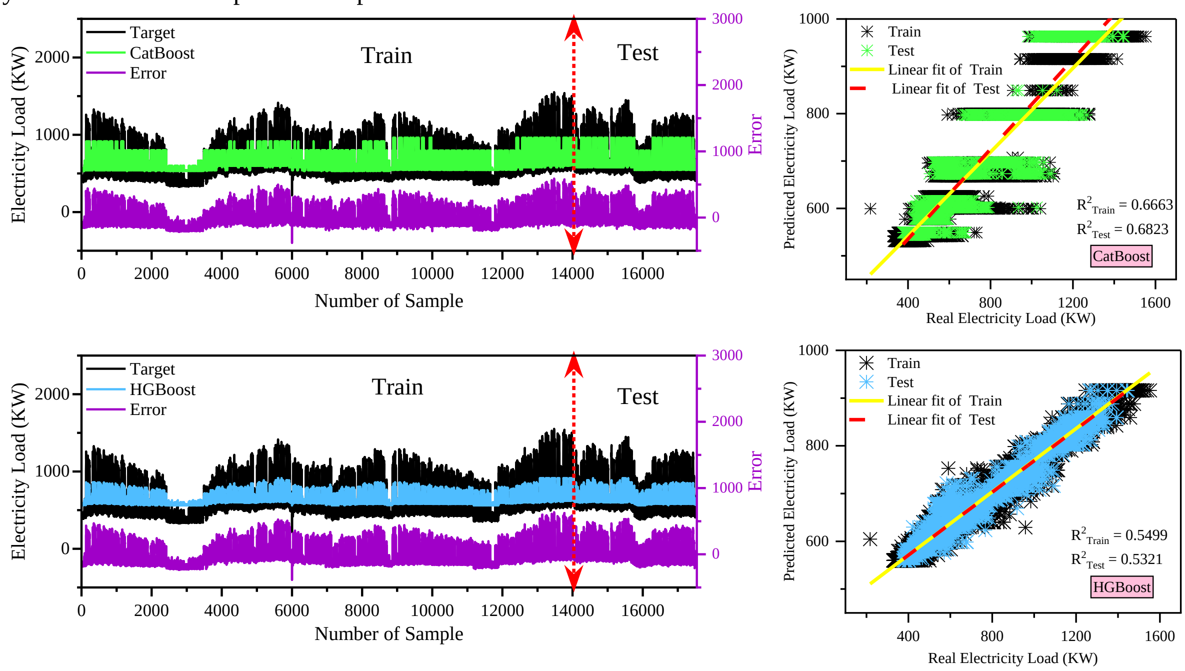

Figure 5: A comprehensive overview of the results obtained from the application of CatBoost and HGBoost models

Table 4: Calculated error metric values for CatBoost and HGBoost models were acquired

Optimizer	CatBoost	HGBoost
	Train	
MBE	-0.09462	1.019514
MAE	116.2694	138.5066
RMSE	145.5828	169.0766
R2	0.66629	0.549892
VAF	66.62898	54.99087
RAE	0.207646	0.241155
	Test	
MBE	-24.0304	-32.2906
MAE	118.7673	143.9044
RMSE	148.3705	180.0433
R2	0.682274	0.532144
VAF	69.06081	54.71933
RAE	0.197646	0.239838

Fig 6 presents a visual representation of the time series data from both training and testing sets for the CatBoost and HGBoost models, enhanced with SSA, ArchOA, and CGO optimizers. The error rates corresponding to each hybrid model are highlighted in purple. Upon analysis, it becomes apparent that the error rate of the testing set is considerably lower for the hybrid HGBoost model compared to other models, indicating the superior predictive capability of the HGBoost hybrids. Notably, among these hybrids, the HGBoost-SSA model demonstrates particularly minimal error. This observation underscores the efficacy of hybridization, particularly when integrating the HGBoost algorithm with the SSA optimization technique, in enhancing predictive precision.

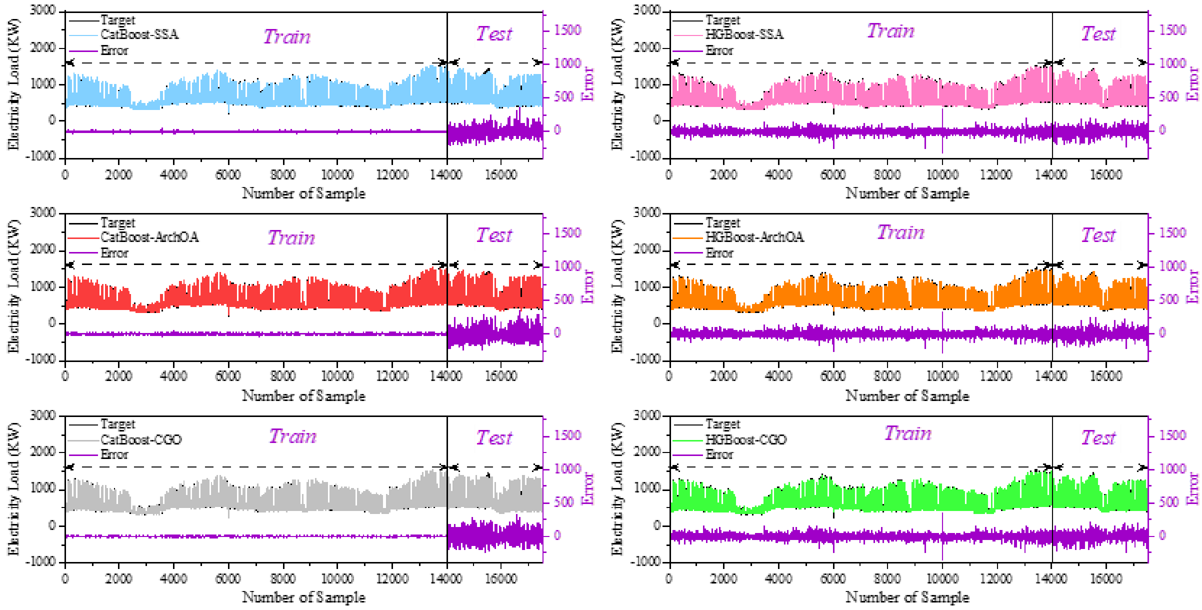

Figure 6: Temporal sequences illustrating the actual and predicted data derived from hybrid models employing CatBoost and HGBoost methodologies

For a thorough investigation of hybrid models, Fig 7 illustrates scatter plots presenting these models alongside the statistical R2 index. Upon detailed scrutiny of the graphical representations, it is evident that the test data within the hybrid HGBoost model exhibits reduced dispersion and a more pronounced clustering along the unity line (x=y) plot. Notably, the HGBoost-SSA model stands out with an impressive R2 value of 0.9802, underscoring its exceptional predictive accuracy. The significant impact of parameter optimization on the effectiveness of HGBoost models must be emphasized. As previously mentioned, the initial R2 value of the HGBoost model was 0.6823. However, as depicted in Fig 6, the refinement of model parameters, coupled with the integration of optimizers, has markedly enhanced the performance of the HGBoost model. Conversely, while there has been some improvement in the performance of other models, the degree of enhancement remains relatively modest. The ability of HGBoost models to yield satisfactory results even with smaller datasets potentially elucidates the superior performance observed in this study among the hybrid HGBoost models. While R^2 serves as a key visual indicator in Fig 7, it is important to note that the overall model evaluation was based on multiple performance metrics, including RMSE, MAE, and RAE, as detailed in Table 4. The HGBoost-SSA model's superiority is not limited to its high R^2 value; it also demonstrated favorable results across these other metrics, indicating lower residual errors and greater generalization. The optimizer's adaptive tuning capability, particularly in handling non-linear variations in the dataset, contributed to this well-rounded performance. This multifaceted evaluation reinforces the robustness of the findings beyond a single statistical index.

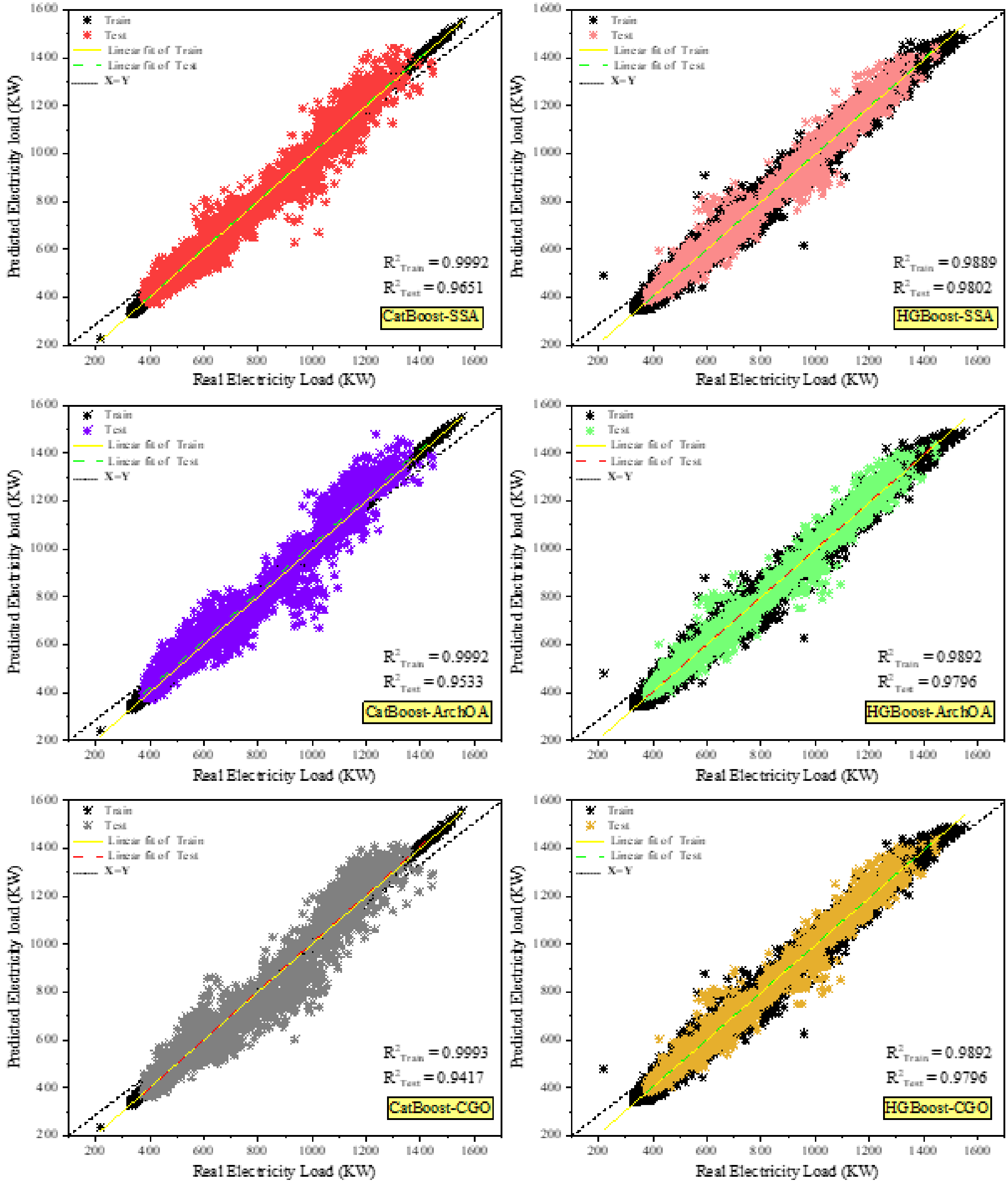

Figure 7: Scatter plot illustrating the alignment between observed and predicted values for hybrid models employing CatBoost and XGBoost algorithms

Fig 8 portrays graphical representations illustrating error metrics associated with the hybrid models. Examination of the R2 and MAE metrics reveals the commendable performance of the hybrid HGBoost models during the prediction phase. Notably, the hybrid HGBoost-SSA model demonstrates superior performance in terms of the R2 metric, recording the highest value among all models, while the CatBoost-CGO model exhibits the lowest R2 value. This consistent performance trend is observed across various other statistical indices, with the HGBoost-SSA model consistently demonstrating superior performance. For a comprehensive and detailed analysis, the corresponding values of these indices for the hybrid models are systematically presented in Table 5.

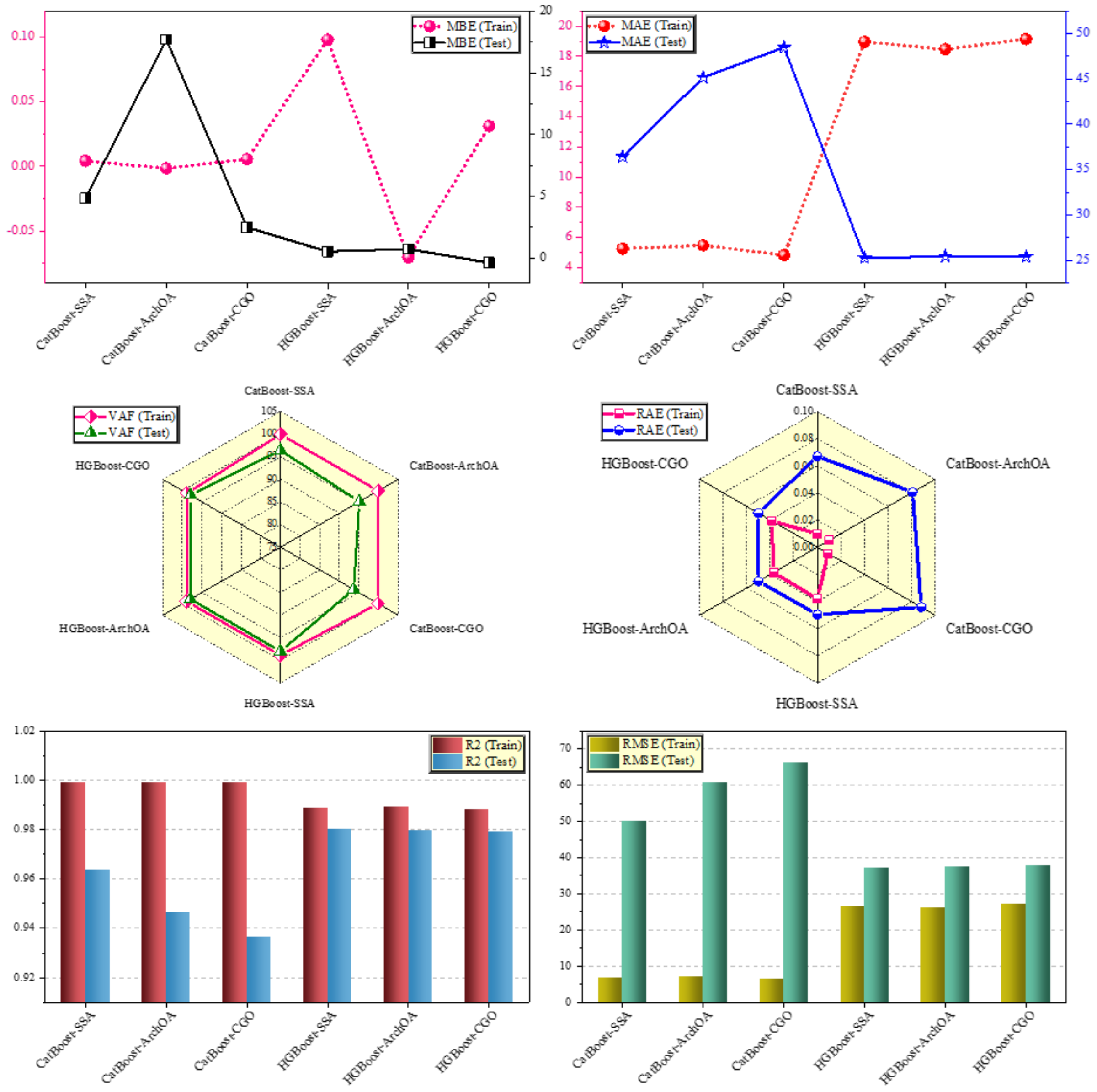

Figure 8: Plots depicting error metrics for the proposed hybrid models

Table 5: Performance indicators derived from the application of CatBoost and HGBoost hybrid models

Optimizer	CatBoost-SSA	CatBoost-ArchOA	CatBoost-CGO	HGBoost-SSA	HGBoost-ArchOA	HGBoost-CGO
`	Train					
MBE	0.00421	-0.0016	0.005508	0.09796	-0.07036	0.031186
MAE	5.256385	5.477229	4.830862	18.96746	18.46572	19.14335
RMSE	6.947521	7.228678	6.411334	26.60791	26.15155	27.06491
R2	0.99924	0.999177	0.999353	0.988853	0.989232	0.988466
VAF	99.924	99.91773	99.93528	98.88528	98.92318	98.84665
RAE	0.009909	0.01031	0.009145	0.037951	0.0373	0.038603
	Test					
MBE	4.840587	17.69604	2.485141	0.531182	0.725461	-0.39068
MAE	36.40991	45.17546	48.49285	25.23514	25.40597	25.35173
RMSE	50.21371	60.76902	66.29697	37.22947	37.65042	37.70935
R2	0.963608	0.946701	0.936563	0.979995	0.97954	0.979476
VAF	96.39464	95.12203	93.66517	97.99994	97.9548	97.94785
RAE	0.06689	0.080951	0.088315	0.049594	0.050155	0.050233

Table 6 summarizes the cross-validation performance of the proposed hybrid models alongside additional baseline models, including LightGBM and XGBoost, under default configurations. The results clearly demonstrate the superior predictive performance of the hybrid models, particularly HGBoost-SSA, which achieved the lowest test RMSE (37.2 kW) and the highest R^2 (0.98). The consistency between test and cross-validation errors indicates the robustness of the proposed approach.

In addition, Table 7 presents the sensitivity analysis results, showcasing the impact of variations in key hyperparameters (learning rate, depth, and number of iterations) on model performance. Among the evaluated models, learning rate emerged as the most sensitive parameter in most configurations, with even small deviations causing noticeable changes in RMSE. These analyses provide further evidence of the stability and reliability of the proposed framework across different parameter settings.

Table 6: Comprehensive model evaluation cross-validation, and baseline comparisons

Model	Test RMSE (kW)	Test MAE (kW)	Test R^2	Optimal Hyperparameters	CV RMSE (Mean ± SD)
HGBoost-SSA	37.2	25.2	0.98	lr=0.085, depth=8, iter=800	37.4 ± 0.8
HGBoost-ArchOA	37.6	25.4	0.979	lr=0.082, depth=9, iter=750	38.0 ± 0.9
HGBoost-CGO	37.7	25.3	0.979	lr=0.088, depth=7, iter=850	38.5 ± 1.1
CatBoost-SSA	50.2	36.4	0.963	lr=0.127, depth=8, iter=743	50.5 ± 1.2
CatBoost-ArchOA	60.7	45.1	0.946	lr=0.115, depth=7, iter=800	58.0 ± 1.3
CatBoost-CGO	66.2	48.4	0.957	lr=0.105, depth=9, iter=700	65.4 ± 1.5
LightGBM	152.6	122.4	0.645	Default parameters	150.6 ± 1.1
XGBoost	165.3	134.7	0.598	Default parameters	164.1 ± 0.9
Base CatBoost	148.4	118.8	0.682	Default parameters	149.2 ± 2.1
Base HGBoost	180	143.9	0.532	Default parameters	181.5 ± 2.8

Table 7: Hyperparameter sensitivity analysis across all models

Model	Learning Rate Range (ΔRMSE%)	Depth Range (ΔRMSE%)	Iterations Range (ΔRMSE%)	Most Sensitive Parameter
HGBoost-SSA	0.06-0.12 (+4.2%)	7-10 (+3.8%)	700-900 (+2.1%)	Learning Rate
HGBoost-ArchOA	0.05-0.11 (+5.1%)	6-9 (+4.5%)	650-850 (+3.3%)	Learning Rate
HGBoost-CGO	0.07-0.13 (+3.9%)	7-10 (+4.1%)	750-950 (+1.9%)	Depth
CatBoost-SSA	0.10-0.15 (+6.7%)	6-9 (+5.9%)	650-850 (+4.8%)	Learning Rate
CatBoost-ArchOA	0.09-0.14 (+7.2%)	5-8 (+6.3%)	600-800 (+5.5%)	Learning Rate
CatBoost-CGO	0.08-0.13 (+6.9%)	7-10 (+5.7%)	700-900 (+4.2%)	Depth
LightGBM	0.08-0.15 (+9.5%)	5-9 (+8.7%)	500-1000 (+6.4%)	Learning Rate
XGBoost	0.07-0.14 (+10.2%)	4-8 (+9.1%)	600-900 (+7.8%)	Learning Rate

Fig 9 illustrates plots presenting the outcomes observed in both the training and testing datasets of hybrid models utilizing CatBoost and HGBoost. In the training dataset, the hybrid CatBoost models display a narrower dispersion and lower error compared to the HGBoost hybrid model, with their median line positioned near zero. Notably, among these variants, the CatBoost-SSA model exhibits superior performance. However, a shift in performance dynamics is noticeable in the testing dataset, where the effectiveness of CatBoost models declines while the hybrid HGBoost models show improved performance. Specifically, the hybrid HGBoost-SSA model stands out by demonstrating reduced dispersion, with its median line closely aligned with zero. These findings imply a decreased error margin, indicating a commendable level of predictive performance.

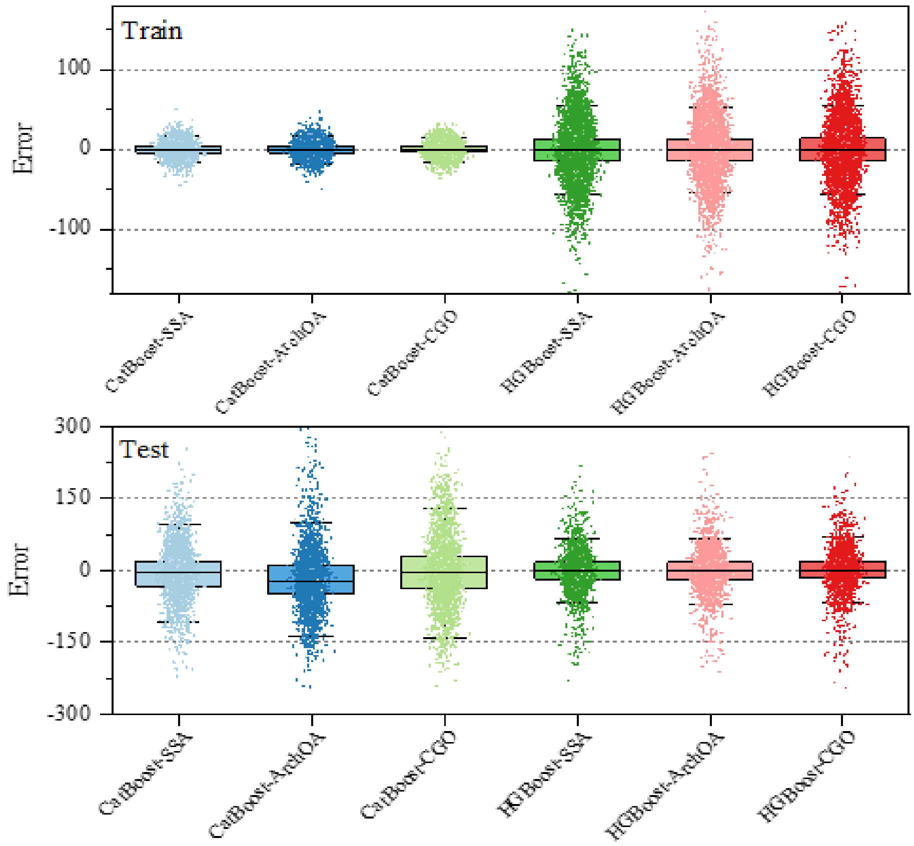

Figure 9: Box plots of error measurements for CatBoost and HGBoost hybrid models during the testing and training phases

In Fig 10, the visual representation illustrates the temporal execution duration of each algorithm over consecutive iterations. The observed fluctuations in execution time suggest that algorithms initially displaying temporal variability generally exhibit lower stability compared to those with oscillations. Nevertheless, as iterations accumulate, these algorithms tend to converge towards a state of increased stability. Notably, the HGBoost-ArchOA and HGBoost-SSA algorithms conform to this observed trend, demonstrating higher stability relative to other algorithms. This stability characteristic functions as an informative indicator of their proximity to an optimal state.

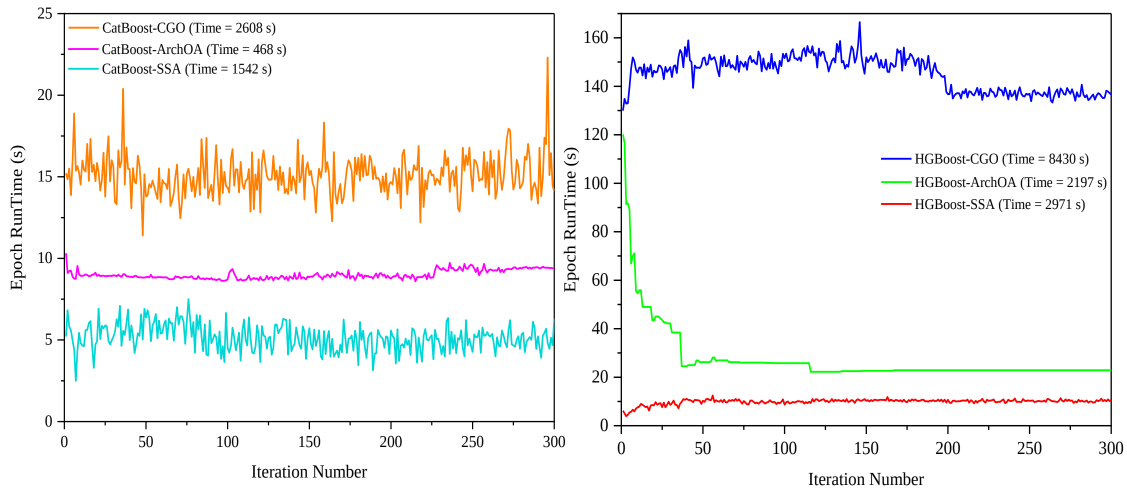

Figure 10: Comparison of runtime for various hybrid models

Fig 11 presents a visual representation depicting the correlation and relative distributions of two metrics, exploitation and exploration, observed over 300 iterations during the exploration phase within hybrid models. The graphical illustration in Fig 11 underscores the rapid convergence of hybrid HGBoost models towards an optimal state. Particularly noteworthy is the observation that, among the models scrutinized, the HGBoost model optimized through the CGO technique displayed the most swift and effective progression toward attaining optimality. Furthermore, notable advancements were also observed in the CatBoost model with CGO optimization.

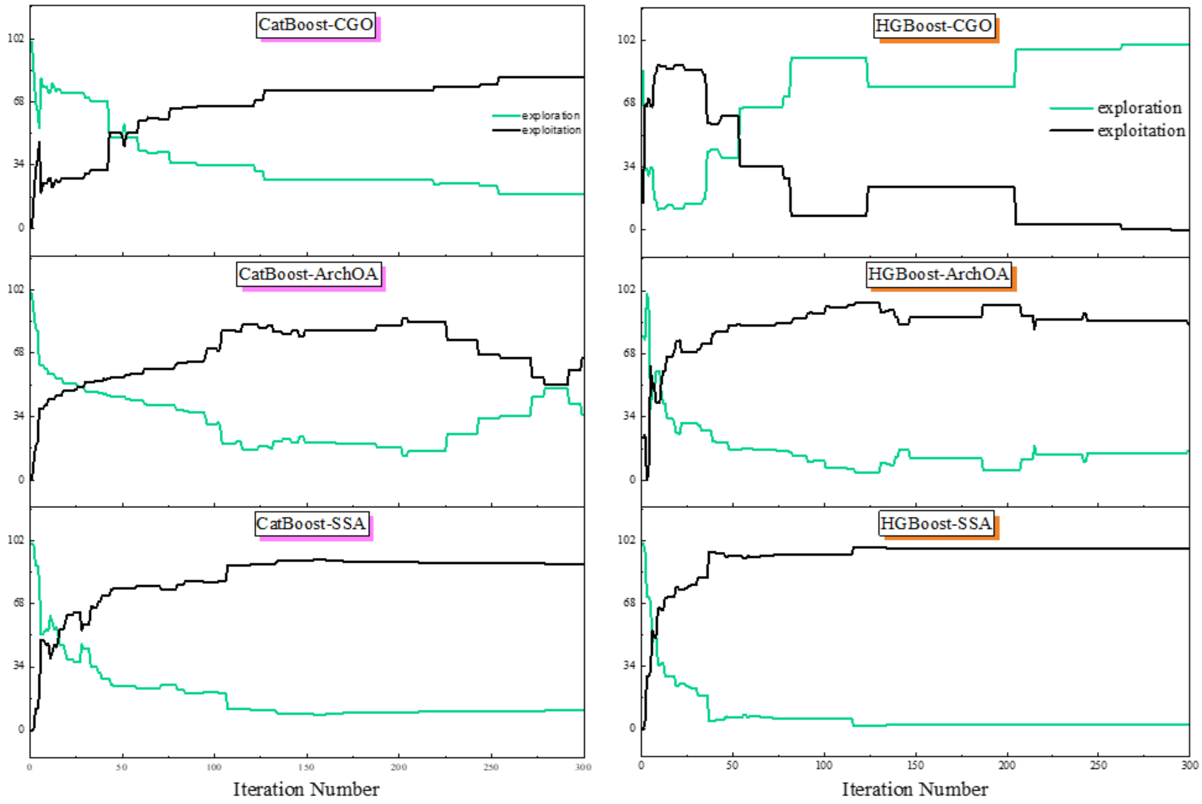

Figure 11: The interplay between exploitation and exploration within the hybrid models

Fig 12 depicts the convergence graph, which highlights the convergence status of hybrid models, with a specific emphasis on CatBoost and HGBoost models. Convergence status is assessed through the Mean Squared Error (MSE) index. It is noteworthy that hybrid HGBoost models demonstrate notably superior convergence performance when compared to other models, as evidenced by considerably lower MSE indices. Within this comparison, all HGBoost models particularly excel, exhibiting the most favorable convergence with the lowest MSE values.

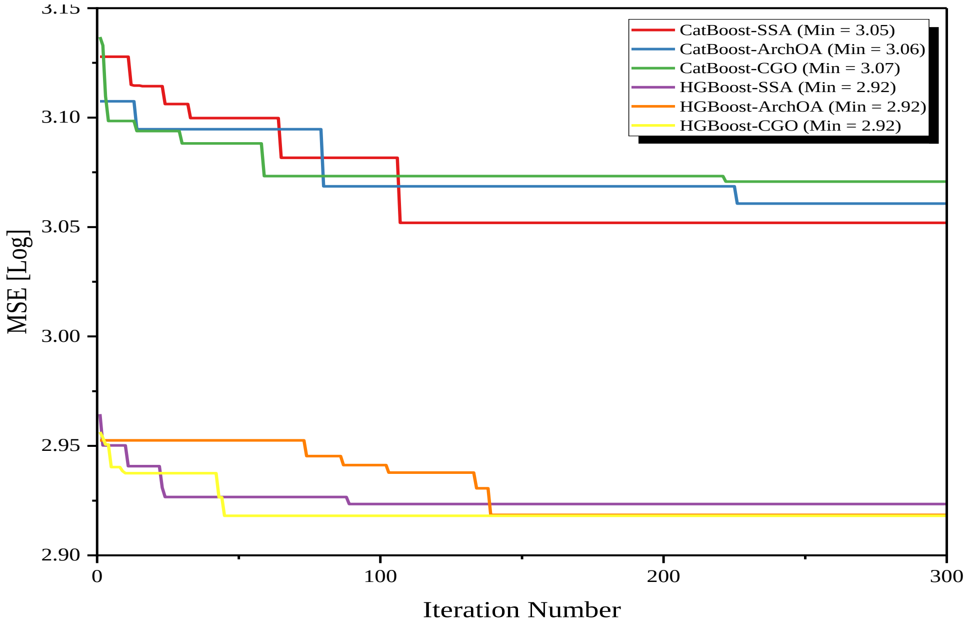

Figure 12: The convergence plots of the CatBoost and HGBoost hybrid models

4 Discussion

This section discusses the performance and implications of the proposed hybrid machine learning models in comparison to existing state-of-the-art methods. The obtained results demonstrate clear advantages, particularly for the CatBoost-CGO model, which achieved an R^2 of 0.9993 on training data and 0.9417 on test data. These values reflect a notable improvement over standalone models such as CatBoost (R^2 = 0.682) and HGBoost (R^2 = 0.532), and also surpass the performance metrics reported in earlier studies, including Random Forest approaches (Ahmad & Chen, 2019) and hybrid ANN-based models (Raza et al., 2017).

This improvement can be attributed to the combined strengths of gradient-boosting techniques and metaheuristic optimization. The use of CatBoost and HGBoost provided a solid foundation due to their ability to handle categorical variables and gradient-based learning efficiently. However, it was the integration of optimization algorithms—specifically CGO, SSA, and ArchOA—that significantly enhanced the models' predictive capabilities. These optimizers allowed for more effective hyperparameter tuning, leading to better convergence, reduced overfitting, and greater accuracy. Among them, the CGO algorithm, when combined with CatBoost, showed the most stable and accurate results across all metrics, as evidenced by the scatter plots and radar charts presented in the results section.

Despite the strong performance, the models are not without limitations. The experiments were conducted using a campus-based dataset from Japan (2020–2021), which may limit the generalizability of findings to broader or more heterogeneous environments. While the dataset includes several influential features statistically analyzed in Table 2 the model's performance in real-time, cross-domain, or data-sparse settings remains untested. To ensure the reproducibility of the proposed approach, all methodological configurations and software specifications have been explicitly detailed. The optimization process employed a population size of 50, a maximum of 200 iterations, and a convergence tolerance set at 1e-6 across all metaheuristic algorithms. The experiments were conducted in a Python 3.9.16 environment utilizing scikit-learn 1.2.2, CatBoost 1.2, HGBoost 0.1.6, numpy 1.24.3, pandas 1.5.3, and optuna 3.1.0 for hyperparameter tuning and model training. These comprehensive specifications are provided to facilitate the replication of results and support future extensions of this research. Additionally, although the use of metaheuristic optimization improved prediction accuracy, it introduced computational complexity that may affect scalability. Future research should evaluate the framework on diverse datasets and explore adaptive learning strategies for real-time applications. Comparisons with deep learning-based hybrid systems could also provide further insight into the relative strengths of the proposed models.

5 Conclusion

In this section, we examine results from individual (CatBoost and HGBoost) and combined modeling approaches (integrating SSA, ArchOA, and CGO). We present outcomes systematically using visual representations and tables. The subsequent sections will analyze these findings, facilitating a thorough exploration of research outcomes. For example, one analysis constructs a correlation matrix, illustrating relationships between input and output variables. Another assesses the significance and sensitivity of input parameters in predicting electricity consumption. Additionally, outcomes from individual models detail superior performance by CatBoost. Visual representations of time series data from hybrid models indicate enhanced predictive capability, particularly HGBoost-SSA. Further analyses elaborate on hybrid model performance, emphasizing the accuracy of certain configurations. Overall, these findings deepen our understanding of predictive capabilities in electricity consumption forecasting.

Nomenclature

acc	Acceleration	MAE	Mean Absolute Error
ArchOA	Archimedes Optimization Algorithm	MBE	Mean Bias Error
CatBoost	Categorical Gradient Boosting	ML	Machine Learning
CGO	Chaos game optimization	MSE	Mean Squared Error
CV	Coefficient of variation	RAE	Relative Absolute Error
den	Density	R^2	Coefficient of Determination
GB	Global Best	rand	Randomly generated numbers within the range of 0 to 1
GBDT	Gradient Boosting Decision Trees	RMSE	Root Mean Square Error
HGBoost	Hybrid Gradient Boosting	SSA	Sparrow search algorithm
HGS	Hunger Games Search	VAF	Variance Accounted For
JSD	Jensen Shannon Divergence	vol	Volume
lb	Lower boundaries	WI	Willmott Index

Competing of interests

The authors declare no competing of interests.

Authorship contribution statement

Peiying Li: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] S. S. Aung, "Electric power is the main driving force for industrialization," 2018.

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

[3] A. Hussain, M. Rahman, and J. A. Memon, "Forecasting electricity consumption in Pakistan: The way forward," *Energy Policy*, vol. 90, pp. 73–80, 2016.

[4] M. Uz Zaman, A. Islam, and N. Sultana, "Short term load forecasting based on internet of things (IoT)," 2018, *BRAC University*.

[5] C. H. J. Kumar and M. Veerakumari, "Load forecasting of Andhra Pradesh grid using PSO, DE algorithms," *Int J Adv Res Comput Eng Technol*, 2012.

[6] T. Van de Graaf, "International energy agency," in *Handbook of governance and security*, Edward Elgar Publishing, 2014, pp. 489–503.

[7] A. Vergnol, "Intégration dans le réseau électrique et le marché de l'électricité de production décentralisée d'origine renouvelable: gestion des congestions locales," 2010, *Ecole centrale de Lille*.

[8] H. Doukas, K. D. Patlitzianas, K. Iatropoulos, and J. Psarras, "Intelligent building energy management system using rule sets," *Build Environ*, vol. 42, no. 10, pp. 3562–3569, 2007.

[9] R. Missaoui, H. Joumaa, S. Ploix, and S. Bacha, "Managing energy smart homes according to energy prices: analysis of a building energy management system," *Energy Build*, vol. 71, pp. 155–167, 2014.

[10] R. E. Abdel-Aal, "Short-term hourly load forecasting using abductive networks," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 164–173, 2004.

[11] Y. Chen *et al.*, "Short-term load forecasting: Similar day-based wavelet neural networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 322–330, 2009.

[12] N. Amjady, F. Keynia, and H. Zareipour, "Short-term load forecast of microgrids by a new bilevel prediction strategyfile:///E:/Article/New Project/222/samplle/scholar (11).ris," *IEEE Trans Smart Grid*, vol. 1, no. 3, pp. 286–294, 2010.

[13] H. Krstić, Ž. Koški, I. I. Otković, and M. Španić, "Application of neural networks in predicting airtightness of residential units," *Energy Build*, vol. 84, pp. 160–168, 2014.

[14] S. Grillo, M. Marinelli, S. Massucco, and F. Silvestro, "Optimal management strategy of a battery-based storage system to improve renewable energy integration in distribution networks," *IEEE Trans Smart Grid*, vol. 3, no. 2, pp. 950–958, 2012.

[15] I. S. Candanedo, E. H. Nieves, S. R. González, M. T. S. Martín, and A. G. Briones, "Machine learning predictive model for industry 4.0," in *Knowledge Management in Organizations: 13th International Conference, KMO 2018, Žilina, Slovakia, August 6–10, 2018, Proceedings 13*, Springer, 2018, pp. 501–510.

[16] A. G. Briones, P. Chamoso, and A. Barriuso, "Review of the main security problems with multi-agent systems used in e-commerce applications," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 5, no. 3, p. 55, 2016.

[17] J. A. Morente-Molinera, G. Kou, R. González-Crespo, J. M. Corchado, and E. Herrera-Viedma, "Solving multi-criteria group decision making problems under environments with a high number of alternatives using fuzzy ontologies and multi-granular linguistic modelling methods," *Knowl Based Syst*, vol. 137, pp. 54–64, 2017.

[18] T. Li, S. Sun, M. Bolić, and J. M. Corchado, "Algorithm design for parallel implementation of the SMC-PHD filter," *Signal Processing*, vol. 119, pp. 115–127, 2016.

[19] J. A. G. Coria, J. A. Castellanos-Garzón, and J. M. Corchado, "Intelligent business processes composition based on multi-agent systems," *Expert Syst Appl*, vol. 41, no. 4, pp. 1189–1205, 2014.

[20] A. González-Briones, J. Prieto, F. De La Prieta, E. Herrera-Viedma, and J. M. Corchado, "Energy optimization using a case-based reasoning strategy," *Sensors*, vol. 18, no. 3, p. 865, 2018.

[21] A. González-Briones, P. Chamoso, H. Yoe, and J. M. Corchado, "GreenVMAS: Virtual organization-based platform for heating greenhouses using waste energy from power plants," *Sensors*, vol. 18, no. 3, p. 861, 2018.

[22] A. González-Briones, P. Chamoso, S. Rodríguez, H. Yoe, and J. M. Corchado, "Reuse of waste energy from power plants in greenhouses through MAS-based architecture," *Wirel Commun Mob Comput*, vol. 2018, 2018.

[23] T. Ahmad and H. Chen, "Utility companies' strategy for short-term energy demand forecasting using machine learning based models," *Sustain Cities Soc*, vol. 39, pp. 401–417, 2018.

[24] T. Ahmad and H. Chen, "Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems," *Sustain Cities Soc*, vol. 45, pp. 460–473, 2019.

[25] T. Saksornchai, W.-J. Lee, K. Methaprayoon, J. R. Liao, and R. J. Ross, "Improve the unit commitment scheduling by using the neural-network-based short-term load forecasting," *IEEE Trans Ind Appl*, vol. 41, no. 1, pp. 169–179, 2005.

[26] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustain Cities Soc*, vol. 35, pp. 257–270, 2017.

[27] D. Liu *et al.*, "Electricity consumption and economic growth nexus in Beijing: A causal analysis of quarterly sectoral data," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2498–2503, 2018.

[28] M. Q. Raza, M. Nadarajah, D. Q. Hung, and Z. Baharudin, "An intelligent hybrid short-term load forecasting model for smart power grids," *Sustain Cities Soc*, vol. 31, pp. 264–275, 2017.

[29] L. Xiao, W. Shao, T. Liang, and C. Wang, "A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting," *Appl Energy*, vol. 167, pp. 135–153, 2016.

[30] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, "Identifying services for short-term load forecasting using data driven models in a Smart City platform," *Sustain Cities Soc*, vol. 28, pp. 108–117, 2017.

[31] P. Pinson and H. Madsen, "Benefits and challenges of electrical demand response: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686–699, 2014.

[32] W.-C. Hong, *Hybrid intelligent technologies in energy demand forecasting*. Springer, 2020.

[33] M. A. AlKhars, "Survey and analysis of the quantitative methods used in electricity research on GCC countries: 1983–2018," *Heliyon*, vol. 5, no. 10, 2019.

[34] X. JIN, "Energy Station Data," 2024, doi: 10.6084/m9.figshare.24978645.v1.

[35] Y. Zhang, Z. Zhao, and J. Zheng, "CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China," *J Hydrol (Amst)*, vol. 588, p. 125087, 2020.

[36] A. Kaushal, A. K. Gupta, and V. K. Sehgal, "Hybrid CatBoost and SVR Model for Earthquake Prediction Using the LANL Earthquake Dataset," *Informatica*, vol. 49, no. 14, 2025.

[37] J. Jiang, B. Cui, C. Zhang, and F. Fu, "DimBoost: Boosting gradient boosting decision tree to higher dimensions," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 1363–1376, 2018, doi: 10.1145/3183713.3196892.

[38] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: sparrow search algorithm," *Systems Science and Control Engineering*, vol. 8, no. 1, pp. 22–34, 2020, doi: 10.1080/21642583.2019.1708830.

[39] L. Zhang, Y. Chen, and Z. Yan, "Predicting the short-term electricity demand based on the weather variables using a hybrid CatBoost-PPSO model," *Journal of Building Engineering*, vol. 71, p. 106432, Jul. 2023, doi: 10.1016/J.JOBE.2023.106432.

[40] E. Dogan and N. Yorukeren, "Enhancement of Transmission System Security with Archimedes Optimization Algorithm," *Turkish Journal of Electrical Power and Energy Systems*, vol. 2, no. 2, pp. 115–123, 2022, doi: 10.5152/tepes.2022.22015.

[41] I. Alsaidan, M. A. M. Shaheen, H. M. Hasanien, M. Alaraj, and A. S. Alnafisah, "Proton exchange membrane fuel cells modeling using chaos game optimization technique," *Sustainability (Switzerland)*, vol. 13, no. 14, pp. 1–24, 2021, doi: 10.3390/su13147911.

[42] A. Rastgoo and H. Khajavi, "A novel study on forecasting the Airfoil self-noise, using a hybrid model based on the combination of CatBoost and Arithmetic Optimization Algorithm," *Expert Syst Appl*, p. 120576, 2023.

[43] O. D. Adeniji, S. O. Adeyemi, and S. A. Ajagbe, "An Improved Bagging Ensemble in Predicting Mental Disorder Using Hybridized Random Forest - Artificial Neural Network Model," *Informatica (Slovenia)*, vol. 46, no. 4, pp. 543–550, Nov. 2022, doi: 10.31449/inf.v46i4.3916.