Volume 18, Issue 4, 182454 (1-10)

Majlesi Journal of Electrical Engineering (MJEE)

https://dx.doi.org/10.57647/j.mjee.2024.1804.54

Short-term electrical load forecasting through optimally configured long short-term memory

Somasundaram Vasudevan*, Kandasamy Jothinathan

Department of Electrical Engineering, Annamalai University, Chidambaram, India.

***Corresponding author:** vasudevanaettneb@gmail.com

Original Research

Received:
4 August 2023
Revised:
10 September 2024
Accepted:
17 September 2024
Published online:
15 December 2024

Abstract:

Short-term electrical load forecasting plays a pivotal role in modern energy systems, addressing the need for accurate predictions of electricity demand within a time frame ranging from a few hours to a few days. Inaccurate predictions can lead not only to operational challenges but also to economic and environmental consequences, highlighting the critical importance of short-term electrical load forecasting in today's energy landscape. This research aims to mitigate these issues by developing an optimally configured Long Short-Term Memory (LSTM) model for short-term electrical load forecasting in Tamil Nadu, specifically targeting the Villupuram region in India. Although LSTM models are known for their effectiveness, achieving optimal performance in short-term load forecasting requires a tailored approach. Hyperparameter optimization is essential for configuring the LSTM model for this purpose, as manual or trial-and-error hyperparameter tuning is time-consuming and computationally intensive. To address this challenge, this research integrates the Cauchy-distributed Harris Hawks Optimization (Cd-HHO) method to optimally configure the LSTM model. The Cd-HHO-optimized LSTM consistently achieves lower Mean Squared Error (*MSE*) than other state-of-the-art methods, with *MSE* values of 0.7225 in the 2017 dataset, 0.974 in the 2018 dataset, and 0.116 in the 2019 dataset.

Keywords: Short-term electrical load forecasting; Long short-term memory; Cauchy-distributed harris hawks optimization; hyperparameters tuning; Uncertainties in weather forecasts; Power system management; Villupuram region

1. Introduction

A stable and uninterrupted power supply is a fundamental requirement for an electrical power generation system. Additionally, these systems rely on "load forecasting" to meet load demand-a process that involves analyzing historical data, behavior, trends, and other factors that impact load to estimate future demand [1]. Accurate power load forecasts enable demand-side management (DSM) approaches to effectively reduce electricity usage [2]. Inaccurate load forecasts can compromise the security and reliability of power systems, potentially leading to imbalanced generation planning, irregular power flows, and system congestion [3]. Moreover, inaccurate load predictions could render efforts to anticipate power demand futile. For power generation resources to produce the necessary amount of energy for continuous and efficient operation, accurate short-term load forecasting is essential [4].

The literature on computational intelligence methodologies for energy load forecasting highlights the complexity of demand patterns, which can be influenced by factors such as climate, season, holidays, weekdays, as well as social activities, economic factors, and power market policies. Over 50 research papers on this topic were reviewed. Since meteorological conditions significantly affect electrical energy demand, these factors must be considered when forecasting demand [5]. Developing a load forecasting model requires incorporating meteorological parameters, particularly due to the influence of high-energy-consumption devices, like air conditioners, on load demand [6].

Traditional statistical algorithms are primarily limited by the assumption of a linear relationship within time series data; as a result, they struggle to predict power load when dealing with irregular, non-linear trends, high noise, and significant fluctuations. Additionally, the prediction error in these systems can increase considerably when the original data changes due to social or environmental influences [7].

While artificial intelligence algorithms can address these issues, they have certain limitations, such as a tendency to fall into local optima, susceptibility to overfitting, and relatively low convergence rates [8]. Neural network algorithms generally have slower learning speeds than traditional statistical methods, as their optimized objective functions are often highly complex. Furthermore, there is a significant risk of training failures in neural networks [9]. Researchers suggest that machine learning and hybrid approaches may be effective tools for managing the non-linear characteristics of load data [10].
The LSTM architecture includes a loop that allows it to store past data for extended periods, making it highly effective for time series data. LSTM has been widely applied to solve various time series challenges, such as well log generation [11], machine translation [12], and natural language processing. It is considered the most effective model for load forecasting, a common time series data problem [13]. To achieve optimal or state-of-the-art results with LSTM networks, several hyperparameters-including the number of recurrent units, network depth, dropout rate, and pre-trained word embeddings-must be selected and optimized. The difference between average and cutting-edge performance is often determined by the choice of hyperparameters [14]. Manual selection based on user experience or traditional random search are the two most common approaches for selecting hyperparameters [15]. However, these methods require extensive trial and error and are time-consuming. An alternative is to use metaheuristic approaches, known for their ability to find near-optimal solutions within a very large search space, to achieve an ideal configuration for energy forecasting models [16].
To compete with other optimizers, this paper proposes a novel, nature-inspired optimization approach. The fundamental concept of the proposed optimizer is inspired by the cooperative hunting tactics of Harris' Hawks, one of nature's most strategic birds, known for pursuing evasive prey (typically rabbits). While traditional HHO can find the global optimal solution, it often requires numerous iterations. The inclusion of Cauchy operators in the proposed approach reduces the number of iterations needed, ensuring an optimal solution in less time than conventional algorithms [17].

2. Literature review

Azeem et al. [18] examine state-of-the-art approaches recently used for electrical load forecasting, highlighting common practices, recent developments, and areas for improvement. Their review explores the procedures, parameters, and relevant sectors considered in load forecasting, providing a detailed analysis of the advantages, disadvantages, and model error rates. To aid researchers in evaluating best practices, the review also emphasizes the distinct features of methods used across residential, commercial, industrial, grid, and off-grid sectors.
Rui Wang et al. [19] propose an electric load forecasting system that addresses the limitations of traditional forecasting models and achieves higher forecasting accuracy than single-model optimization. This system integrates various classical forecasting methods, hybrid optimization algorithms, and data preparation techniques. Test results indicate that the proposed hybrid model can more accurately represent real values and has been effectively used for dispatch in smart grids.
Umar Javed et al. [20] provide a comprehensive analysis of contemporary linear and non-linear parametric modeling methodologies aimed at ensuring stable and reliable power system operations by addressing non-linearities in electrical load data. Temporal and climatic characteristics are identified as potential input factors in various modeling approaches, based on the results of exploratory data analysis. Several advanced linear and non-linear parametric techniques are evaluated for accuracy using real-time electrical load and climatic data from Lahore, Pakistan.
To predict short-term load demand in a typical microgrid, Usman Bashir Tayab et al. [21] developed a hybrid solution that combines the best-basis stationary wavelet packet transform with a Harris Hawk optimization-based feed-forward neural network. The basis and weights of the neurons in the feed-forward neural network are optimized using Harris Hawk optimization. The proposed model, when compared to other existing models, shows potential for estimating load demand in the Queensland electric market.
Ibrahim Yazici et al. [22] proposed a method for short-term load forecasting based on Video Pixel Networks and one-dimensional Convolutional Neural Networks (CNNs). Although CNNs are less commonly used in time series forecasting compared to traditional models like LSTM and GRU, they exhibit exceptional effectiveness in pattern extraction. In this study, these models are employed to provide forecasts for the next hour and the next 24 hours using actual power usage data. The proposed one-dimensional CNN model outperforms others, with statistical tests showing a mean absolute percentage error of 2.21% for 24-hour predictions. The CNN approach also performs well for 1-hour forecasts, achieving a mean absolute percentage error of approximately 1%. This study provides valuable insights that can enhance prediction models used in the electrical industry.
Ke Li et al. [23] introduce the CEEMDAN-1SE-LSTM model for ultra-short-term power load forecasting in Changsha, China, which is crucial for power dispatch and market development. By decomposing and reconstructing power load data while considering meteorological and holiday factors, the model outperforms others, achieving *RMSE*, MAE, and *MAPE* values of 62.102, 47.490, and 1.649%, respectively. The research significantly enhances forecasting accuracy, supports power dispatch in Changsha, and serves as a valuable reference for other cities developing similar models.
Due to the significant impact of climate change on power usage and emerging trends in smart networks, medium- and long-term electrical load forecasting has become increasingly important. Navid Shirzadi et al. [24] investigated this issue to compare and improve district-level models for predicting electrical load demand. They used a variety of techniques, including LSTM recurrent neural networks, random forests, non-linear auto-regressive exogenous neural

networks, support vector machines, and deep learning approaches. The results show that the deep learning model provided more accurate load demand predictions than SVM and RF, achieving an R-squared value of approximately $0.93-0.96$ and a mean absolute percentage error of about $4-10\%$.

Wendi Li et al. [25] present HELP (Hyperparameter Exploration LSTM-Predictor), a random exploration approach that combines probability-based exploration with LSTM-based prediction. HELP is more efficient in discovering superior hyperparameters with less time spent. Based on experimental data, HELP identifies hyperparameters for both convolutional neural networks and generative adversarial networks, yielding better outcomes and faster convergence.

3. Proposed methodology

Demand response management is a key component that helps reduce peak load and variation in electrical load, and it depends on accurate load forecasting, making it essential to all aspects of electric utility operations. Accurate electrical load forecasting is crucial for managing the power system and energy dispatch, which are vital for the successful operation of the country's economy and the daily lives of its people. On the other hand, poor power load predictions are ineffective. An inaccurate forecast can lead to abnormal power system operation, causing power outages in affected areas and ultimately resulting in significant losses and disasters. This research aims to create a load forecasting model using LSTM. To improve performance, it incorporates hyperparameter adjustment, which is crucial for effective training. However, a random search for hyperparameters can be time-consuming and may take a while to converge due to the need to train deep neural networks with numerous parameters for each selected hyperparameter. The novelty of this research lies in the integration of optimization techniques to identify the optimal hyperparameters for LSTM. Specifically, it combines the Cauchy distribution as a parallel updating strategy with traditional HHO to determine the optimal hyperparameters for LSTM.

3.1 Dataset description

In our research, we use real-time datasets of electrical load consumption and weather information collected from a 230/110 KV auto substation in the Villupuram region of Tamil Nadu, India. From 2017 to 2019, 96 samples per day were taken from the aggregated electrical load data (representing the entire Villupuram region), recorded at 15-minute intervals. The meteorological dataset includes information on wind speed, rainfall, temperature, and humidity, which was obtained from an official Indian government online database portal and recorded at 15-minute intervals between 2017 and 2019. This study incorporates eight independent variables: temperature, wind speed, humidity, rainfall, groundwater level, day type, special days, and the season of the corresponding year. Due to the influence of climatic factors (wind speed, temperature, humidity, and rainfall) on load demand, there is a strong interdependence between these variables and thermal inertia. In turn, the power system load demand is significantly affected by climatic factors. Therefore, considering these factors is essential for developing an accurate short-term load forecasting model.

3.2 Recurrent Neural Network (RNN)

RNNs are extensively used for classification tasks. Gradient descent backpropagation can be employed to train the network with new information. However, a disadvantage of conventional training algorithms is their slow convergence rate, making it difficult to find the global minimum of the error function, as gradient descent often gets trapped in local minima. To optimize complex problems, nature-inspired meta-heuristic algorithms offer derivative-free solutions. Supervised machine learning techniques, which account for all dependent features, enable classification with higher accuracy and reduce the misclassification rate in load forecasting, which can occur as frequently as 20% of the time. According to Fausett (1994), an RNN is an ANN with arbitrary connections between neurons, typically fully connected between neighbouring layers. The hidden state values of the hidden layer in the previous state $h\nu(t_m\text{-}1)$ and the current data point $x\,(t_m)$ are the inputs that the network nodes receive. As a result, by virtue of recurrent connections, inputs at time t_m have an effect on the network's future outputs. A standard RNN with an input vector $I\nu = (I\nu_1,...,I\nu_T)$ calculates a hidden vector $h\nu = (h\nu_1,...,h\nu_T)$ and an output vector $O\nu = (O\nu_1,...,O\nu_T)$ by iterating equation (1) and (2) over $t_m = 1, \ldots, T$.

$$h\nu(t_m) = A_F(Wt_{(hx)}x^{(t_m)} + Wt_{(hh)}h\nu^{(t_m-1)} + bl_h) \quad (1)$$

$$y\nu^{(t_m)} = \sigma(Wt_{(hy)}h\nu^{(t_m)} + bl_y) \quad (2)$$

where: bl_y and bl_h are vectors of biases, $Wt_{(hx)}$, $Wt_{(hh)}$ and $Wt_{(hy)}$ are weights matrices of the input-hidden layer, hidden-output layer and recurrent connections respectively. An activation function is A_F [24]. Conventional neural networks are trained utilizing the back propagation over time algorithm across a number of time steps.

3.3 Long term short term memory networks

RNNs from the 1980s are the foundation for LSTM networks. To address the vanishing and exploding gradient issues in conventional RNNs, Hochreiter and Schmidhuber developed the LSTM architecture [22]. Their design allows information to be retained during processing and uses feedback to track prior network states. Compared to traditional RNNs, the LSTM model has shown exceptional ability to learn long-term dependencies in real-world applications. As a result, LSTM models are widely used in cutting-edge applications. Typically, the LSTM consists of several memory blocks, each containing memory cells and gates. The gates control the flow of information, while the memory cells retain the network's temporal state through self-connections. Each memory block contains an input gate, an output gate, and a forget gate. The output gate regulates how cell activations are distributed throughout the network. This study uses supervised learning procedures, with 80% of the real-time monitoring database allocated for training and 20% for testing. The architecture of the LSTM cell is shown in Figs. 1 and 2.

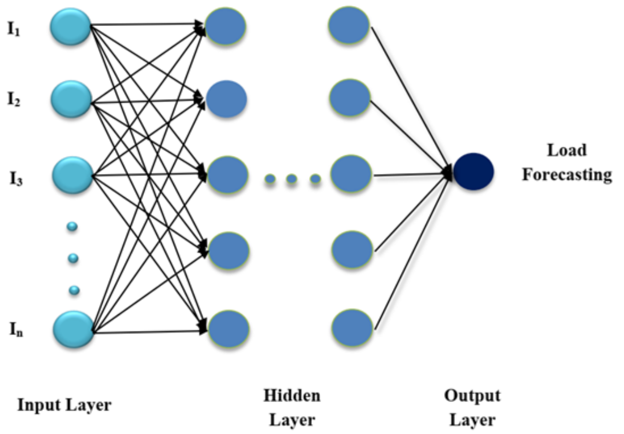

Figure 1. Flow diagram of LSTM in Load Forecasting.

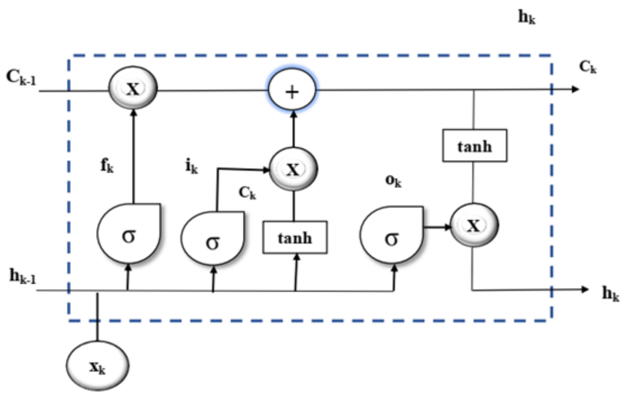

Figure 2. RNN-LSTM layer model.

3.4 Cauchy distributed-Harris hawks optimization (Cd-HHO)

This strategy is inspired by the prey investigation, surprise pounce, and various attacking techniques used by Harris hawks. With the correct formulation, it can be applied to solve any optimization problem, as it is population-based and gradient-free. The primary objective is to enhance electrical load forecasting performance by optimizing the LSTM classifier's hyperparameters using the Cd-HHO algorithm. The research framework considers factors such as batch size and the number of hidden neurons. The Cd-HHO technique aims to minimize the Root Mean Square Error ($RMSE$) of the electrical load forecasting model iteratively, stopping once the halting criteria are met. It begins with an initial set of solutions (the hyperparameters), which are randomly generated.

3.4.1 Exploration phase

HHO is inspired by the behavior of Harris' hawks, which can track and identify their prey using their keen vision, even when the prey is difficult to spot. The hawks wait, observe, and scan the arid terrain in search of prey, sometimes for many hours. In HHO, the Harris' hawks, which represent potential solutions, are evaluated at each stage to determine if they are the expected prey or close to it. The hawks in HHO employ one of two strategies to wait for prey: they perch at random locations. If we assume the same probability p for each perching strategy, they either perch based on the positions of other family members (so they can attack the rabbit nearby), as shown in equation (3) for the condition of $p < 0.5$, or they perch on random tall trees (random locations inside the group's home range), as shown in equation (3) for the condition of $p \geq 0.5$.

$$S(i+1) = \begin{cases} S_{rand}(i) - rm_1|S_{rand}(i) - 2rm_2S(i)| & p \geq 0.5 \\ (S_{rabbit}(i) - S_m(i)) - \\ \quad rm_3(lb + rm_4(ub - lb)) & p < 0.5 \end{cases} \tag{3}$$

where rm_1, rm_2, rm_3, rm_4, and p are random values inside (0,1), which are updated in each iteration; $S(i+1)$ is the position vector of the hawks in the next iteration; $S_{rabbit}(i)$ is the position of the rabbit; $S(i)$ is the current position vectors of the hawks; and ub and lb. $S_{rand}(i)$ is a hawk randomly chosen from the current population, and S_m is the average position of the current population of hawks. Show the upper and lower bounds of the variables. Using the equation (4), the hawks' average location is determined.

$$S_m(i) = \frac{1}{n}\sum_{i=1}^{n} S_l(i) \tag{4}$$

where $S_l(i)$ denotes the positions of each hawk in iteration i and n is the total number of hawks. There are a number of ways to determine the average location, but we used the simplest rule.

3.4.2 Fitness computation

LSTM networks serve as the fitness function, evaluating and returning the electrical load forecasting results with the lowest $RMSE$. The fitness function is computed using equation (5). Once the minimum objective function is reached and the corresponding hyperparameters are adjusted, the process is optimized. The separation, alignment, and cohesion coefficients are calculated to update the position and velocity of the agents following the fitness evaluation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}} \tag{5}$$

3.4.3 Exploration to exploitation transition

The Cd-HHO algorithm can switch between exploratory and exploitative behaviors depending on the prey's ability to escape. As the prey attempts to flee, it loses a significant amount of energy. To simulate this, equation (6) is used to model the prey's energy.

$$EN = 2EN_0\left(1 - \frac{i}{MI}\right) \tag{6}$$

Where EN denotes the prey's escape energy, the highest amount of iterations is MI, and EN_0 represents the prey's energy at its beginning state.

In Cd-HHO, EN_0 fluctuates at random during each repetition within the range (-1, 1). When EN_0's value falls from 0 to -1, the rabbit is physically weakened; when it rises from 0 to 1, the rabbit is physically strengthened. Throughout the iterations, there is a diminishing tendency in the dynamic escape energy EN. As a result, the suggested technique performs the exploration phase, and when $|EN| \geq 1$, the algorithm attempts to take advantage of the solutions' close proximity during the exploitation phases. When the escape energy $|EN| < 1$, the hawks scout out potential rabbit hiding places in various locations. In other words, when $|EN| < 1$, exploitation occurs, and when $|EN| \geq 1$, exploration occurs.

3.4.4 Exploitation phase

In the initial phase, Harris' hawks carry out a surprise pounce on their intended prey. However, the prey often attempts to escape dangerous situations. As a result, various pursuit strategies emerge in real-life scenarios. Four potential ways to model the attacking phase are proposed in the Cd-HHO, based on the prey's escape behavior and the hawks' pursuit tactics. The prey continually tries to evade perilous situations. Assume PSE represents the likelihood that a prey would either escape (PSE < 0.5) or not successfully flee (PSE ≥ 0.5) before a surprise pounce. Regardless of what the prey does, the hawks will perform either a forceful or gentle siege to capture it. Depending on the amount of energy the prey has stored, the hawks will encircle it from different directions, either softly or strongly. In fact, the hawks approach their prey slowly to increase their chances of delivering a sudden, fatal strike. Over time, as the prey becomes exhausted, the hawks will intensify their siege strategy to quickly capture the weakened victim. This method is imitated by the EN option, which enables the Cd-HHO to switch between soft and strong besiege strategies. The soft besiege takes place in this case when $|EN| \geq 0.5$, but the severe besiege occurs when $|EN| < 0.5$.

3.4.5 Soft besiege

The rabbit still has enough energy when PSE ≥ 0.5 and $|EN| \geq 0.5$ and it tries to flee by making a series of erroneous jumps at random, but ultimately is unable to. The Harris' hawks gradually circle the rabbit in these attempts to exhaust it before making the surprise pounce. Equation (7) - Equation (8) following rules serve as a model for this behaviour:

$$S(i+1) = \Delta S(i) - E|jS_{rabbit}(i) - S(i)| \tag{7}$$

$$\Delta S(i) = S_{rabbit}(i) - S(i) \tag{8}$$

Where $j = 2(1 - r5)$ is the random jump strength of the rabbit during the fleeing process, $r5$ is a random value inside (0, 1), and $\Delta S(i)$ is the dissimilar between the rabbit's position vector and the current location in iteration i. The j number changes randomly with each cycle to mimic the erratic character of rabbit locomotion.

3.4.6 Hard besiege

When PSE ≥ 0.5 and $|EN| < 0.5$, the prey is so exhausted and it has a low escaping energy Furthermore, the Harris' hawks hardly circle their chosen prey before making their final surprise pounce. In this case, the equation (9) is utilized to update the current positions.

$$S(i+1) = S_{rabbit}(i) - EN|\Delta S(i)| \tag{9}$$

3.4.7 Soft besiege with progressive rapid dives

The rabbit has sufficient energy to effectively flee while $|EN| \geq 0.5$ but PSE < 0.5, and a mild besieged is still present before the surprise pounce. Compared to the previous approach, this method is more logical. Levy flights are employed in the Cd-HHO method to mathematically describe the leapfrog movements and prey-escaping patterns. These flights mimic the erratic, sudden, and rapid dives hawks make around escaping prey, as well as the zigzagging, deceptive movements that prey animals use when fleeing. Hawks conduct multiple quick, coordinated dives around their prey while gradually adjusting their position and direction in response to the prey's evasive maneuvers. Observations of competitive behaviors in nature further support this mechanism. In non-destructive foraging conditions, these flight-based behaviors have been shown to be the most effective searching strategies for predators and foragers. Additionally, LF-based patterns have been observed in the chasing behaviors of animals like sharks and monkeys. Consequently, Levy flight-based motions are utilized in this stage of the Cd-HHO approach.

Based on actual hawk behavior, we hypothesized that hawks gradually select the best dive when attempting to catch prey in a competitive context. Therefore, we proposed that hawks may assess (or decide) their next course of action based on the following rule in equation (10) in order to carry out a modest siege:

$$NM = S_{rabbit}(i) - EN|jS_{rabbit}(i) - S(i)| \tag{10}$$

Then, to determine whether the dive will be successful, they compare the potential outcome of the movement with that of the previous dive. If the outcome seems unlikely, they begin to dive irregularly, abruptly, and quickly as they approach the rabbit, especially when they observe that the prey is becoming more cunning. The following rule in equation (11) was used to model their dives based on levy flight patterns.

$$FM = NM + rv^{*}LF(ps) \tag{11}$$

Equation (12) is used to determine this, where ps is the problem's dimension, rv is a random vector of size $1 \times ps$, and LF is the Levy flight function.

$$LF(s) = 0.01 * \frac{k1 * \sigma}{|k2|^{\frac{1}{\beta}}}; \sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\Pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}} \right)^{\left(\frac{1}{\beta}\right)} \tag{12}$$

Where β is a default constant set to 1.5 and $k1$ and $k2$ are random numbers inside of (0,1). Equation (13), then, can be used to establish the ultimate plan for updating the hawks' positions during the gentle besiege phase.

$$S(i+1) = \begin{cases} NM & \text{if } F(NM) < F(S(i)) \\ FM & \text{if } F(FM) < F(S(i)) \end{cases} \tag{13}$$

where NM and FM are obtained using equation (10) and (11).

To be clear, during some cycles, this image also records and displays the position history of LF-based leapfrog movement patterns. The coloured dots represent the LF-based patterns' location footprints in one trial before the Cd-HHO arrives at the FM site. Only the best position NM or FM will be chosen as the next location in each stage. All search agents are used in this method.

3.4.8 Hard besiege with progressive rapid dives

The rabbit has the energy to escape when $|EN| < 0.5$ and PSE < 0.5, resulting in the construction of a hard besiege until the use of a surprise pounce to capture and kill the prey. In this step, the prey is placed on the same side of the hawks as it was in the gentle besiege, but this time the hawks are making an effort to shorten the gap between their usual location and the fleeing prey. Due to the hard besiege condition shown in equation (14), the subsequent rule is put into effect.

$$S(i+1) = \begin{cases} NM & \text{if } F(NM) < F(S(i)) \\ FM & \text{if } F(FM) < F(S(i)) \end{cases} \tag{14}$$

Where, NM and FM are obtained using new rules in equation (15) and (16).

$$NM = S_{rabbit}(i) - EN|jS_{rabbit}(i) - S_m(i)| \tag{15}$$

$$FM = NM + rv \times LF(ps) \tag{16}$$

where $S_m(i)$ is obtained using equation (4).

3.4.9 Cauchy distribution

The Cauchy distribution, also known as the Lorentz distribution, is a continuous probability distribution. By refining the accuracy of the candidate solution to within a few percent of the global solution during each iteration of the evolution process, it is possible to enhance the searchability of an optimization process. Initially generated candidate solutions can then move to better positions as the process progresses. In this research, the Cauchy distribution method is modified at the end of the updating process, in addition to conventional updating methods, to further improve searchability and convergence speed.

The Cauchy distribution function is defined in the following equation (17):

$$y = \frac{1}{\pi} \arctan\left(\frac{\gamma}{g}\right) + \frac{1}{2} \tag{17}$$

According to the following equation (18), the corresponding density function is described,

$$f_{Cauchy(0,g)} = \frac{1}{\pi}\left(\frac{g}{g^2+\gamma^2}\right) \tag{18}$$

where, g = 1 is the scaling parameter, y is a number uniformly distributed between [0, 1], and γ = tan (π(y - 1 /2)). Therefore, the Cauchy distribution is an effectively operating mutation applied to enhance searchability and convergence speed [26].

3.4.10 Termination

The process of this method will terminate once an optimal solution is found based on the objective function or fitness function. If not, the algorithm will continue. Additionally, the operation of the algorithm will be suspended once the allowed number of iterations is exceeded. Moreover, if the algorithm is unable to find the fitness value of the best solution after many iterations, it will be discontinued.

4. Result and discussion

Electrical load forecasting is a crucial step in ensuring optimal and profitable electricity market operation. The developed model is evaluated based on performance metrics and compared to other single and hybrid deep learning models. For the purpose of load prediction, the research uses MATLAB 2021a. By improving scheduling and management practices, load forecasting aims to minimize the difference between generation and demand, as well as reduce electricity losses. The performance of the proposed method is assessed using Mean Squared Error (MSE), $RMSE$, Skewness, Mean Absolute Percentage Error ($MAPE$), and Coefficient of Variation (CV).

Figs. 3 to 5 (Tables 1 to 3) below illustrate how the model's performance compares to actual values. It is evident that the proposed method outperforms other comparable strategies for the 2017, 2018, and 2019 datasets in terms of achieving the closest load predictions to actual values. The results highlight the importance of fine-tuning the LSTM's hyperparameters for load forecasting and demonstrate that the proposed technique outperforms conventional algorithms. Integrating the Cauchy distribution in the updating process enhances the searchability and convergence speed of the traditional HHO algorithm, enabling candidate solutions to move toward better positions during this process.

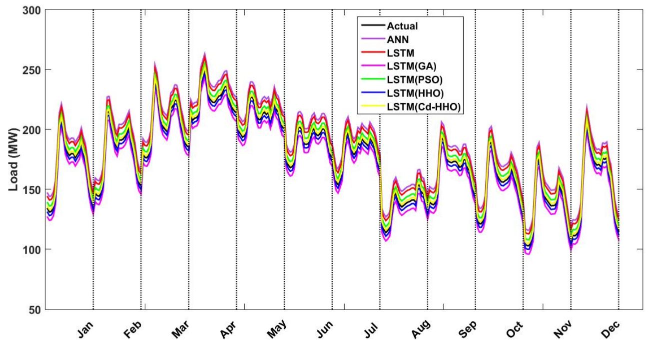

Figure 3. Actual vs technique wise model performance for the year 2017.

Table 1. Evaluation of different models for predicting short-term electrical load in 2017.

Techniques	MSE	RMSE	MAPE	Skew	CV
ANN [20]	10.3041	3.21	1.905	0.000932	5.797
LSTM	6.0025	2.45	1.454	0.000414	3.391
LSTM-GA	4.5369	2.13	1.264	-0.00027	2.631
LSTM-PSO	3.4225	1.85	1.098	0.000178	1.940
LSTM-HHO	1.5625	1.25	0.742	-5.50E-05	0.902
LSTM-Cd-HHO	0.7225	0.85	0.504	1.73E-05	0.412

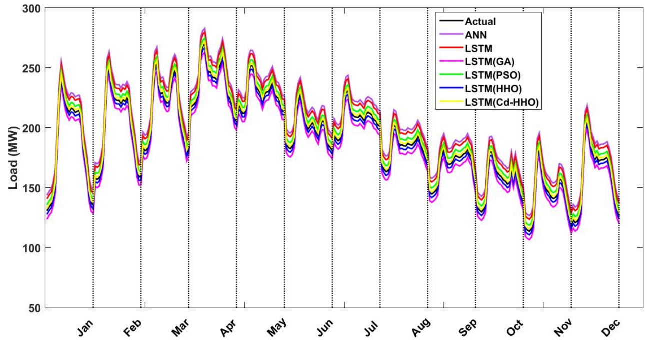

Figure 4. Actual vs technique wise model performance for the year 2018.

Table 2. Evaluation of different models for predicting short-term electrical load in 2018.

Techniques	MSE	RMSE	MAPE	Skew	CV
ANN [20]	8.703	2.95	1.609	0.00057	4.516
LSTM	5.382	2.32	1.265	-0.000276	2.871
LSTM-GA	3.803	1.95	1.063	0.000164	1.983
LSTM-PSO	2.126	1.458	0.795	-6.86E-05	1.129
LSTM-HHO	1.020	1.01	0.551	2.28E-05	0.535
LSTM-Cd-HHO	0.974	0.987	0.538	2.13E-05	0.511

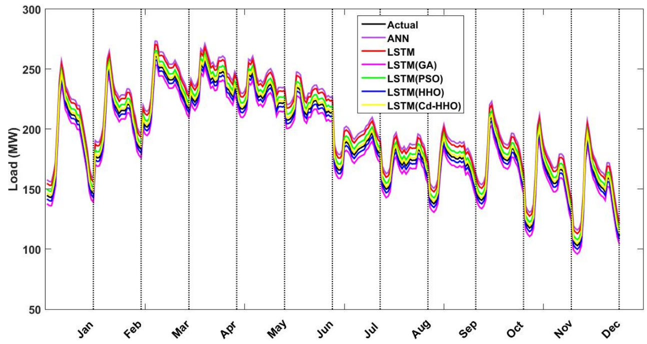

Figure 5. Actual vs technique wise model performance for the year 2019.

Table 3. Evaluation of various models for predicting short-term electrical load in 2019.

Techniques	MSE	RMSE	MAPE	Skew	CV
ANN [20]	7.508	2.74	1.483	-0.00043	3.977
LSTM	4.623	2.15	1.164	-0.00021	2.441
LSTM-GA	3.764	1.94	1.050	0.000154	1.946
LSTM-PSO	2.205	1.485	0.804	-6.89E-05	1.160
LSTM-HHO	0.888	0.942	0.510	1.76E-05	0.461
LSTM-Cd-HHO	0.116	0.34	0.184	8.27E-07	0.060

4.1 Mean Square Error (*MSE*)

MSE is calculated utilizing the average, or more precisely the mean, of errors squared from data relating to a function. *MSE* measures how close the model forecasted electrical load values to a set of actual load values. In the following equations (19) to (23) A_i and P_i are the actual and the forecasted values through the models (employed techniques) correspondingly. Whereas n refers the number of times the summation iteration happens.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|P_i - A_i|^2 \tag{19}$$

The comparison of *MSE* values for various techniques in short-term electrical load forecasting across the years 2017, 2018, and 2019 reveals distinctive performance patterns. In 2017, ANN exhibited the highest *MSE* at 10.3041, while LSTM-GA, LSTM-PSO, LSTM-HHO, and LSTM-Cd-HHO progressively achieved lower *MSE* values, with LSTM-Cd-HHO showcasing the best performance at 0.7225. Similarly, in 2018 and 2019, LSTM-Cd-HHO consistently outperformed other methods, demonstrating significantly lower *MSE* values (0.974 and 0.116, respectively). Notably, traditional ANN displayed higher *MSE* values compared to LSTM-based techniques, and among the LSTM variants, LSTM-Cd-HHO consistently exhibited superior accuracy in short-term load forecasting. These results underscore the efficacy of LSTM-Cd-HHO as a technique that consistently achieves lower prediction errors across multiple years, making it a promising approach for enhancing the accuracy of short-term electrical load forecasting.

4.2 Root Mean Square Error (*RMSE*)

RMSE, sometimes referred to as root mean square deviation, is one of the most frequently used methods to assess the accuracy of forecasts. It measures the deviation of the model's predicted load values from the actual load values using Euclidean distance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}} \tag{20}$$

Examining the *RMSE* values for various techniques employed in short-term electrical load forecasting over the years 2017, 2018, and 2019 reveals distinctive trends. In 2017, ANN exhibited the highest *RMSE* at 3.21, while LSTM-GA, LSTM-PSO, LSTM-HHO, and LSTM-Cd-HHO progressively achieved lower *RMSE* values, with LSTM-Cd-HHO demonstrating the best performance at 0.85. Similarly, in 2018 and 2019, LSTM-Cd-HHO consistently outperformed other methods, showcasing significantly lower *RMSE* values (0.85 and 0.34, respectively). The traditional ANN consistently displayed higher *RMSE* values compared to LSTM-based techniques. Among the LSTM variants, LSTM-Cd-HHO consistently exhibited superior accuracy in short-term load forecasting, indicating its effectiveness in minimizing prediction errors across the three years. These results emphasize the reliability of LSTM-Cd-HHO as a technique that consistently achieves lower *RMSE* values, highlighting its potential for enhancing the accuracy of short-term electrical load forecasting models.

4.3 Mean Absolute Percentage Error (*MAPE*)

One of the most popular methods for evaluating forecast accuracy is the *MAPE* statistic, which offers the advantages of scale independence and interpretability. It is the average of the percentage errors, reflecting how accurately the forecasted load values compare with the actual load values.

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{P_i - A_i}{A_i}\right|}{n} \times 100 \tag{21}$$

Analyzing the *MAPE* values for various techniques utilized in short-term electrical load forecasting across the years 2017, 2018, and 2019 reveals notable trends. In 2017, ANN exhibited the highest *MAPE* at 1.905, while LSTM-Cd-HHO achieved the lowest at 0.504. Similarly, in 2018 and 2019, LSTM-Cd-HHO consistently outperformed other methods, displaying substantially lower *MAPE* values (0.538 and 0.184, respectively). Traditional ANN consistently demonstrated higher *MAPE* values in comparison to LSTM-based techniques. Among the LSTM variants, LSTM-Cd-HHO consistently exhibited superior accuracy in short-term load forecasting, showcasing its efficacy in minimizing percentage errors across the three years. These results emphasize the reliability of LSTM-Cd-HHO as a technique consistently yielding lower *MAPE* values, underlining its potential to enhance the precision of short-term electrical load forecasting models.

4.4 Coefficient of Variance (*CV*)

The coefficient of variation (*CV*), also known as the standard deviation to mean ratio, indicates the degree of variance relative to the population mean. The dispersion increases as the *CV* rises.

$$CV = \frac{\sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n-1}}}{\bar{A}} \times 100 \tag{22}$$

Analyzing the *CV* values for various techniques employed in short-term electrical load forecasting over the years 2017, 2018, and 2019 reveals important insights into the stability and consistency of each method's forecasting errors. In 2017, ANN exhibited the highest *CV* at 5.797, indicating a higher relative variability in errors, while LSTM-Cd-HHO demonstrated the lowest *CV* at 0.412, signifying a more stable and consistent performance. Similarly, in 2018 and 2019, LSTM-Cd-HHO consistently outperformed other methods, displaying substantially lower *CV* values (0.511 and 0.06, respectively). The decreasing trend in *CV* values across the years for LSTM-Cd-HHO suggests a consistent improvement in stability and reliability. Traditional ANN consistently showed higher *CV* values compared to LSTM-based techniques. Among the LSTM variants, LSTM-Cd-HHO consistently exhibited superior stability in short-term load forecasting, emphasizing its effectiveness in providing more consistent and reliable predictions.

4.5 Skew

The *skew* error is a high-dimensional distribution error that shows the distribution of predicted values relative to actual values. If its value lies between -0.5 and 0.5, this range is considered acceptable.

$$Skew = \frac{\sum_{i=1}^{n}(P_i - A_i)^3/n}{SD^3} \qquad (23)$$

where *SD* refers the standard distribution
Analyzing the Skewness values for various techniques employed in short-term electrical load forecasting over the years 2017, 2018, and 2019 provides insights into the distribution characteristics of forecasting errors. In 2017, ANN demonstrated a positive skewness of 0.000932, indicating a slightly right-skewed distribution, while LSTM-Cd-HHO showed a smaller positive skewness of 1.73E-05, suggesting a nearly symmetrical distribution. Similarly, in 2018 and 2019, LSTM-Cd-HHO consistently displayed skewness values close to zero, indicating a relatively symmetric distribution of errors. The other techniques, including LSTM, LSTM-GA, LSTM-PSO, and LSTM-HHO, also exhibited skewness values around zero, suggesting a balanced distribution of errors over the three years. These skewness values offer insights into the shape of the error distribution for each technique, contributing to a comprehensive understanding of their forecasting performance.

5. Conclusion

This research addresses key challenges in short-term electrical load forecasting, a critical aspect of modern energy management. It emphasizes the importance of accurate short-term predictions, ranging from a few hours to a few days, due to their operational, economic, and environmental impacts. Focused on Villupuram, Tamil Nadu, India, the study developed a customized LSTM model for forecasting regional electricity demand. Recognizing that standard LSTM configurations can be inefficient for this purpose, the research employed hyperparameter optimization, specifically using the Cauchy-distributed Cd-HHO approach, to effectively fine-tune the model. This innovative optimization significantly improved forecasting accuracy, achieving lower *MSE*, *RMSE*, *MAPE*, and *CV* scores over three years (2017 – 2019) compared to other advanced methods. The optimized LSTM model not only enhances academic knowledge but also provides a practical and precise solution for real-world short-term electrical load prediction applications.

Authors contributions
All authors have contributed equally to prepare the paper.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interests
The authors reported no potential conflicts of interest regarding the research, writing, and publication of this paper.

References

[1] Y. Liang, D. Niu, and W.C. Hong. "**Short term load forecasting based on feature extraction and improved general regression neural network model**". *Energy*, 166:653–663, 2019. DOI: https://doi.org/10.1016/j.energy.2018.10.119.

[2] Z. Guo, K. Zhou, X. Zhang, and S. Yang. "**A deep learning model for short-term power load and probability density forecasting**". *Energy*, 160:1186–1200, 2018. DOI: https://doi.org/10.1016/j.energy.2018.07.090.

[3] K.S. Khan, S.M. Ali, Z. Ullah, I. Sami, B. Khan, C.A. Mehmood, U. Farid, J. Saleem, and J.S. Ro. "**Statistical energy information and analysis of Pakistan economic corridor based on strengths, availabilities, and future roadmap**". *IEEE Access*, 8:169701–169739, 2020. DOI: https://doi.org/10.1109/ACCESS.2020.3023647.

[4] E.K. Lee, W. Shi, R. Gadh, and W. Kim. "**Design and implementation of a microgrid energy management system**". *Sustainability*, 8(11):1143, 2016. DOI: https://doi.org/10.3390/su8111143.

[5] S.N. Fallah, R.C. Deo, M. Shojafar, M. Conti, and S. Shamshirband. "**Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directions**". *Energies*, 11(3):596, 2018. DOI: https://doi.org/10.3390/en11030596.

[6] Y. Kim, H.G. Son, and S. Kim. "**Short term electricity load forecasting for institutional buildings**". *Energy Rep*, 5:1270–1280, 2019. DOI: https://doi.org/10.1016/j.egyr.2019.08.086.

[7] J. Song, J. Wang, and H. Lu. "**A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting**". *Appl. Energy*, 215:643–658, 2019. DOI: https://doi.org/10.1016/j.apenergy.2018.02.070.

[8] X. Zhao, C. Wang, J. Su, and J. Wang. "**Research and application based on the swarm intelligence algorithm and artificial intelligence for wind farm decision system**". *Renew. Energ*, 134:681–697, 2019. DOI: https://doi.org/10.1016/j.renene.2018.11.061.

[9] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu. "**Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting**". *Appl. Energy*, 180:213–233, 2016. DOI: https://doi.org/10.1016/j.apenergy.2016.07.113.

[10] E. Ceperic, V.Ceperic, and A. Baric. "**A strategy for short-term load forecasting by support vector regression machines**". *IEEE Trans. Power Syst*, 28(4):4356–4364, 2013. DOI: https://doi.org/10.1109/TPWRS.2013.2269803.

[11] D. Zhang, C.H.E.N. Yuntian, and M.E.N.G. Jin. "**Synthetic well logs generation via Recurrent Neural Networks**". *Pet. Explor. Dev*, 45(4):629–639, 2018. DOI: https://doi.org/10.1016/S1876-3804(18)30068-5.

[12] Y. Chen and D. Zhang. "**Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory**". *Adv. Appl*, 1:100004, 2021. DOI: https://doi.org/10.1016/j.adapen.2020.100004.

[13] B.S. Kwon, R.J. Park, and K.B. Song. "**Short-term load forecasting based on deep neural networks using LSTM layer**". *J. Electr. Eng. Technol*, 15:1501–1509, 2020. DOI: https://doi.org/10.1007/s42835-020-00424-7.

[14] N. Reimers and I. Gurevych. "**Optimal hyperparameters for deep lstm-networks for sequence labeling tasks**". *arXiv preprint arXiv*, 1707.06799, 2017. DOI: https://doi.org/10.48550/arXiv.1707.06799.

[15] J. Bergstra and Y. Bengio. "**Random search for hyper-parameter optimization**". *J. Mach. Learn. Res*, 13(2), 2012. URL https://www.jmlr.org/papers/v13/bergstra12a.html.

[16] S. Bouktif, A. Fiaz, A. Ouni, and M.A. Serhani. "**Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting**". *Energies*, 13(2):391, 2020. DOI: https://doi.org/10.3390/en13020391.

[17] Z.A. Ali, H. Zhangang, and W.B. Hang. "**Cooperative path planning of multiple UAVs by using max-min ant colony optimization along with cauchy mutant operator**". *Fluct. Noise Lett*, 20(01):2150002, 2021. DOI: https://doi.org/10.1142/S0219477521500024.

[18] A. Azeem, I. Ismail, S.M. Jameel, and V.R. Harindran. "**Electrical load forecasting models for different generation modalities: a review**". *IEEE Access*, 9(142239-142263), 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3120731.

[19] R. Wang, J. Wang, and Y. Xu. "**A novel combined model based on hybrid optimization algorithm for electrical load forecasting**". *Appl. Soft Comput*, 82:105548, 2019. DOI: https://doi.org/10.1016/j.asoc.2019.105548.

[20] U. Javed, K. Ijaz, M. Jawad, E.A. Ansari, N. Shabbir, L. Kütt, and O. Husev. "**Exploratory data analysis based short-term electrical load forecasting: A comprehensive analysis**". *Energies*, 14(17):5510, 2021. DOI: https://doi.org/10.3390/en14175510.

[21] U.B. Tayab, A. Zia, F. Yang, J. Lu, and M. Kashif. "**Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform**". *Energy*, 203:117857, 2020. DOI: https://doi.org/10.1016/j.energy.2020.117857.

[22] I. Yazici, O.F. Beyca, and D. Delen. "**Deep-learning-based short-term electricity load forecasting: A real case application**". *Eng. Appl. Artif. Intell.*, 109:104645, 2022. DOI: https://doi.org/10.1016/j.engappai.2021.104645.

[23] K. Li, W.Huang, G. Hu, and J. Li. "**Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network**". *Energy and Buildings*, 279:112666, 2023. DOI: https://doi.org/10.1016/j.enbuild.2022.112666.

[24] N. Shirzadi, A. Nizami, M. Khazen, and M. Nik-Bakht. "**Medium-term regional electricity load forecasting through machine learning and deep learning**". *Designs*, 5(2):27, 2021. DOI: https://doi.org/10.3390/designs5020027.

[25] W. Li, W.W. Ng, T. Wang, M. Pelillo, and S. Kwong. "**HELP: An LSTM-based approach to hyperparameter exploration in neural network learning**". *Neurocomputing*, 442:161–172, 2021. DOI: https://doi.org/10.1016/j.neucom.2020.12.133.

[26] L. Liu, D. Zhao, F. Yu, A.A. Heidari, C. Li, J. Ouyang, H. Chen, M. Mafarja, H. Turabieh, and J. Pan. "**Ant colony optimization with Cauchy and greedy Levy mutations for multilevel COVID 19 X-ray image segmentation**". *Comput. Biol. Med.*, 136:104609, 2021. DOI: https://doi.org/10.1016/j.compbiomed.2021.104609.