SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I3P103

Volume 13 Issue 3, 23-37, March 2026
© 2026 Seventh Sense Research Group®

Original Article

Adaptive Hybrid Swallow Search Optimized Hybrid Deep Learning Model for Energy Demand Forecasting

G.S. Bibin[1*], H. Vennila[2], M Chinchu[3]

[1,2]*Department of EEE, Noorul Islam Centre for Higher Education, Tamil Nadu, India.*
[3]*Department of ECE, College of Engineering, Kidangoor, Kerala, India.*

[*]*Corresponding Author : bibings81@gmail.com*

Received: 05 December 2025 Revised: 05 January 2026 Accepted: 04 February 2026 Published: 31 March 2026

Abstract *- Energy consumption has become a critical challenge due to the environmental and economic implications of the current technological development. Artificial Intelligence techniques are used increasingly for optimal electricity demand forecasting and maintaining efficient energy management. However, existing AI-based models often fail to capture complex nonlinear demand patterns because they focus on short-term dependencies, which limits their ability to adapt to Dynamic Energy Consumption (EC) patterns. To overcome these drawbacks, a novel Energy Demand Forecasting using Optimized Resilient serial Cascaded Encoder-LSTM (EDFORCE) technique has been proposed in this paper. The proposed method necessitates historical data on electricity consumption and forecasts electricity demand based on season, day, and time interval. The proposed model uses a Gaussian Mixture Modelling (GMM) based data clustering to obtain season-based segmented data for demand modeling. Further, Attention-Enhanced serial Cascaded Encoder Long Short Term Memory (A-SCE-LSTM) networks are trained to forecast electricity demand, which is optimized by using the novel Hybrid Swallow Search-Adam Optimization (SWO-Adam) algorithm. The efficiency of the EDFORCE approach has been assessed using measures such as Accuracy (AC), Precision (PR), Recall (RE), F1-score (F1-s), Relative Root Mean Square Error (rRMSE), Mean Absolute Percentage Error (MAPE), Mean Bias Error (MBE), EC, and computation time. The proposed EDFORCE model achieves a higher accuracy of 97%, whereas the previous models, such as K-PCD, FNET, and VW AA-KELM model, achieve accuracy of 95%, 93%, and 92%. Additionally, the proposed EDFORCE model consumed 23.8%, 14.9%, and 1.3% less energy than the K-PCD, FNET, and VW AA-KELM models, respectively.*

Keywords *- Energy Demand Forecasting, Gaussian Mixture Modelling, Swallow Search Optimization Algorithm, Serial Cascaded Encoder Long Short Term Memory Network.*

1. Introduction

Energy consumption is influenced by the activities and behaviors of energy consumers, which are driven by their preferences [1]. Consumer preferences are likely to fluctuate over time, introducing uncertainty into the daily EC pattern [2]. Economic conditions, climatic change, holidays, working days, time periods, and social and behavioral aspects are other factors that impact the consumption of energy [3]. Furthermore, power demand is increasing due to rising population and the introduction of several domestic appliances such as dishwashers and clothes dryers [4]. According to a report by Powercor Australia [5], climate change is another factor contributing to the rise in EC. It causes a spike in energy demand for about 0.5% or 40 hours annually.

Increasing electrical assets, such as generation, distribution, and transmission infrastructure, to handle the surge in demand is one strategy [6]. However, this asset growth is inappropriate because the benefits of fulfilling the demand spike, which only occurs 0.5% of the time, are substantially exceeded by the expense of establishing the electrical infrastructure [7]. Energy management techniques used on the demand side are another method of controlling the sudden rise in energy usage [8, 9]. Accurate electrical energy demand forecasting is necessary for managing peak EC [10].

Conventional electric power forecasting techniques lower the uncertainty by grouping customers with similar patterns into smaller groups or combining the total power use of a large number of households [11]. However, it is crucial to accurately forecast the household's electric power usage during the smart grid construction process in order to create a personalized electricity pricing plan for that specific household [12]. Furthermore, since the majority of the dependent variables, including electric current, weather, voltage, etc., are unknown, univariate data prediction continues to be one of the most challenging machine learning issues [13–15]. Traditional univariate forecasting techniques

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

are typically used in situations where there are either too many variables to measure or the remaining features are too complex to measure, such as stock market index forecasting difficulties [16]. These univariate forecasting techniques are made more flexible while requiring no additional data [17]. As long as the management system has access to past data, these methods can be used to forecast the power usage of other homes [18]. Deep Learning (DL) models have gained popularity globally in recent years. They are widely utilized in many different application domains, including Natural Language Processing (NLP), image object detection, and time series analysis, which were also utilized for energy demand prediction [19].

Even though significant advancements have been made with the help of deep learning-based forecasting models, the majority of existing methods are still limited by certain factors [20]. Most of the DL models are unable to accurately predict the demand in the electricity consumption data during various times due to the presence of non-linear, multi-scale temporal dependencies, and also the real-world sudden demand fluctuations. Moreover, the majority of existing works either rely on dataset-specific tuning or use very expensive computational model architectures, thereby limiting their scalability and deployment to practical smart-grid environments. Thus, there is a very clear research gap in the development of a forecasting framework for energy demand that would be able to learn complex temporal behavior, stay computationally efficient, and change according to the season and time-dependent variations in demand all at once. To overcome these drawbacks, a novel Energy Demand Forecasting using Optimized Resilient serial Cascaded Encoder-LSTM (EDFORCE) technique has been proposed in this paper. The proposed EDFORCE framework presents a novel forecasting approach through its unique combination of techniques, which outperforms previous forecasting models that either use traditional LSTM architectures or very costly hybrid networks. The slack of the proposed combination consists of (i) season-aware probabilistic clustering with GMM, (ii) an attention-enhanced serial cascaded encoder-LSTM for multi-scale temporal learning, and (iii) a mixed Swallow-Search-Adam optimization technique for stable hyper-parameter tuning. Thus, the effective demand forecasting with less computational effort and better accuracy has been developed by the integration of the aforementioned methods, which has not been achieved by the previously reported methods. The significant contributions of the EDFORCE model are as follows.

- The primary goal of the proposed EDFORCE approach is to develop an effective energy demand forecasting technique for efficiently capturing the nonlinear electricity spikes and accurately forecasting the demand.
- The proposed approach uses the historic data and predicts electricity demand for the season, day, and time interval by integrating a clustering-based effective deep learning model.
- The GMM-based clustering is performed to attain season-based segmented data for extracting the complicated relationships in the historical data, and helps in enhancing the prediction.
- Additionally, the novel A-SCE-LSTM model has been employed to forecast electricity demand by using the hybrid SWO-Adam model for effectively tuning model parameters. The SWO-Adam model combines the benefits of SWO and Adam optimizers to make them capture more complex temporal dependencies.
- The efficacy of the developed EDFORCE approach has been assessed utilizing measures such as AC, PR, RE, F1-s, MBE, MAPE, and rRMSE.

The remainder of the proposed EDFORCE model has been provided as follows. Section 2 discusses the literature review for energy demand forecasting. Section 3 describes the proposed EDFORCE model in detail. Section 4 discusses the experimental evaluation of the proposed EDFORCE model, and Section 5 discusses the conclusion and future work.

2. Literature Review

With the growing EC patterns in the modern era, effective energy management is required to meet energy needs and reduce energy waste. Various studies were conducted to determine the most accurate, reliable, and fastest technique for forecasting the energy demand. This section discusses some of the relevant contributions in the energy demand forecasting domain.

2.1. Statistical and Early Forecasting Approaches

Initial attempts to predict electricity demand mainly relied on classical time-series and statistical learning methods. In 2024, Feng et al. [21] developed an EC forecast approach based on LSTM and Transformer models that thoroughly consider vehicle and environmental aspects, paying particular attention to individual driving patterns and driving situations. After considering each driver's unique driving style and the driving conditions of both types of vehicles, the experiment demonstrates that the model's MAPE is decreased by 18.47% and 15.27%, respectively.

In 2024, Sulaiman et al [22] proposed using Kolmogorov-Arnold Networks (KAN) to simulate complex nonlinear dynamics with great accuracy. Based on real data from a commercial building, the research introduced KAN as a unique application for real-world chiller energy prediction. According to the findings, KAN performed better than TLBO-DL (R2: 0.9366, RMSE: 6.2892) and NN (R2: 0.9281, RMSE: 6.7709).

In 2025, Hussain et al. [23] introduced deep learning frameworks that used Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithms to forecast electricity consumption in Bangladesh's western area. Wintertime enhances forecast performance, with average RMSE and MAPE decreases of 14.61 MW and 1.53%, respectively, as a result of continuously decreasing load demand.

Though the adoption of these techniques to some extent reduced the prediction error over the linear models, still these methods were not so effective in handling the highly complicated loads, and also suffer from uncertainty due to the change in consumers' behavior and the intricate time component that is involved in the real-world consumption data.

2.2. Clustering and Optimization-Based Approaches

The technique of clustering-driven load segmentation is gaining popularity in identifying different consumer groups according to their seasonal behavior and trends. In 2025, Yang et al [24] presented a novel K-PCD technique that was well-suited for creating EC time series. K-PCD uses a Distance Measure (PCD) based on the Pearson correlation coefficient and a novel centroid computation technique that considers both the PCD and the conventional Euclidean Distance (ED) between time series. The results show that K-PCD outperformed typical clustering techniques.

In 2025, Ahmed et al [28] presented a framework for optimal energy consumption allocation prediction, employing the Least Squares Support Vector Machine (LSSVM), which was optimized by the new Gooseneck Barnacle (GB) Optimizer and Selective Opposition-based constrained BMO. According to the findings, the new GB base-optimized LSSVM offered a solid and dependable resolution for daily EC allocation prediction with an accuracy of 99.98%.

Clustering and hybrid optimization approaches not only help in increasing the interpretability of load patterns but also in improving the overall precision of demand modelling. However, most of the segmentation techniques are based on deterministic clustering and are rarely integrated end-to-end within deep learning-based forecasting pipelines.

2.3. Deep Learning Models for Energy Demand Forecasting

As deep learning models have been developed, many researchers have tried adopting Neural-Network-Based forecasting architectures to depict the electricity demand dynamics of high Complexity. In 2025, Pai et al. [25] devised a TinyML-optimized multivariate LSTM approach that relies on self-supervised learning to enhance forecast AC with little computing cost. Empirical results showed significant improvement over traditional forecasting models, with MSE of 0.02063 and RMSE of 0.14363, MAE of 0.107, MAPE of 0.155, and R^2 of 0.724.

In 2025, Binbusayyis, A., and Sha, M. [26] presented an innovative city system that leverages Deep Learning Techniques and the Household EC dataset. Regression uses the Modified Deep CNN-Bi-LSTM with attention method to extract complex temporal and spatial data. Regression results from analysis conducted to evaluate improved prediction accuracy using evaluation metrics including MSE, MAPE, and RMSE, which establishes the system's effectiveness. In particular, the suggested model obtains an MAPE of 324.12, MAE of 0.22, and MSE of 0.123.

In 2025, Zhao et al. [27] presented a multimodal information fusion method that combines textual and time series data to take advantage of complementary information from many sources. The CNN and Bi-GRU architectures have been combined to create a hybrid predictive model. The suggested method is superior in both short- and long-term forecasting tasks on the same dataset, as shown by comparative trials with ARIMA, standalone GRU, and EEMD-ARIMA.

Deep learning models have been able to demonstrate better performance; however, there are still many challenges ahead. For instance, existing models struggle to capture multi-scale temporal behavior, seasonal consumption variability, and abrupt demand changes. Besides, model robustness and generalization are still weighed down by dataset characteristics and tuning strategies.

2.4. Specialized Energy Demand Forecasting Frameworks

Conventional Deep Learning and Clustering-Based Forecasting methods are not the only approaches used for such purposes, and besides these, the specialized forecasting frameworks are primarily concerned with the aspects of interpretability, robustness of modeling, and domain-specific applicability in the prediction of energy demand.

In 2025, Ghimire et al. [29] devised a novel explainable electricity demand Forecasting Technique Called Deeply-Fused Nets (FNET). The model integrated climate variables significantly to better understand AC and EC patterns and outputs point forecasts and confidence-interval estimates for daily demand. In the last step, the FNET framework used residual bootstrapping, which involves the resampling of the residuals that were computed as the difference between the actual and predicted values to form the 95% prediction intervals.

In 2025, Wang, Y., and Yang, X. [30] developed a new Vector Weighted Average Kernel Extreme Learning Machine (VW AA-KELM) approach to improve EC forecasting in cloud computing systems. Experimental results show that VWAA-KELM outperforms other models, with a coefficient of determination (R2) of 0.987 in the training set and R2 = 0.973 in the test set, indicating excellent generalization. These specialized modeling techniques indicate that energy analytics are increasingly focusing on explainability and custom forecasting design; on the other hand, they do not directly make use of the combination of probabilistic seasonal segmentation, deep temporal learning, and hybrid optimization.

According to the literature review provided, a majority of the studies have been conducted on certain datasets or have been based on specific regional energy profiles, which means that their adaptability and scalability to larger or real-time industrial environments are very limited. The present-day techniques are not able to provide a consistent integration of probabilistic seasonal segmentation with nonlinear temporal modeling in a single framework. The computational resources utilized by the models were often pervasive, or the Complexity of the architectures makes the deployment difficult in resource-constrained systems. To overcome these drawbacks, a novel EDFORCE technique has been proposed and discussed in the next section.

3. Proposed Method

In this section, a novel Energy Demand Forecasting using Optimized Resilient serial Cascaded Encoder-LSTM (EDFORCE) technique has been proposed, which uses an effective deep learning model for reducing complexity in prediction.

The proposed EDFORCE method makes use of past electricity usage data to predict the demand for the different seasons, days, and specific time intervals. Initially, the dataset is preprocessed and divided into different segments according to the seasons through the use of the Gaussian Mixture Model (GMM) clustering technique. Once this segmentation is completed, the load trend analysis is performed to create the metadata at the interval level, which is then used to aid the demand prediction process. Further, a novel Attention-Enhanced Serial Cascaded Encoder LSTM approach has been introduced to forecast electricity demand, which is optimized by employing the Hybrid Swallow Search-Adam Optimization Algorithm. The Hybrid SWO-Adam Algorithm integrates the benefits of the Adam Algorithm with the social behaviour of birds. Figure 1 shows the overall block diagram of the developed EDFORCE approach.

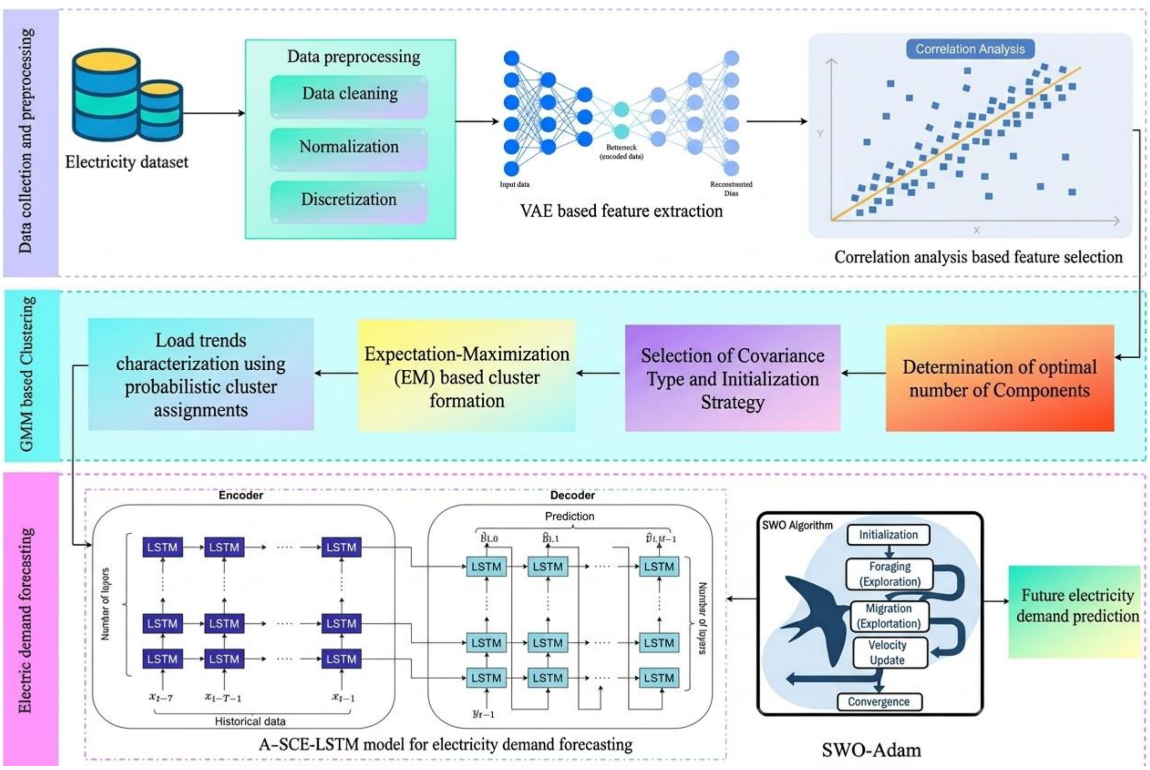


Fig. 1 EDFORCE framework

3.1. Dataset Collection

The UT Chandigarh per day (Timestamp) energy demand database(http://www.moef.gov.in/sites/default/files/Chandigarh) is used in the suggested model. The UT Chandigarh is the capital of Punjab and Haryana and a well-planned metropolis in northwest India, close to the Shivalik hill range of the Himalayas. The city's land area is around 114.10 m × 62. Since January 2013, UT's electricity consumption data has been taken at regular intervals of 15 minutes for a lustrum. Block numbers are assigned at each 15-minute period, ranging from Block 1 (00:00 am to 00:15 am) to Block 96 (11:45 pm to 00:00 am).

3.2. Preprocessing

The real-world dataset can be vulnerable to several inconsistencies, such as incomplete data, missing values, raw format, noise, etc. These inconsistencies or mistakes in the raw data could result in inadequate data evaluation. In order to guarantee the rationality of the procedure of knowledge discovery from the data, the collected data must be pre-processed. Data cleaning comprises filling in missing numbers, removing noise, detecting outliers, and resolving disputes in the data. A preprocessing technique called data normalization resizes the range of feature values to a particular scale, typically between 0 and 1. Binning idea hierarchies are used in discretization to prepare data for analysis.

3.3. Feature Extraction

The features from the preprocessed data have been extracted using the Variational Autoencoder (VAE) technique. Auto Encoders (AEs) are trained to reconstruct their input data, which serves as the foundation for VAEs. Nevertheless, new samples cannot be produced by conventional autoencoders. In order to overcome this restriction, VAEs impose a variational constraint by requiring that the latent variable z have a normal distribution. The VAE-based feature extraction process is represented as follows:

$$M(y,\widehat{y}) = ||x-\hat{x}|| \tag{1}$$

$$log\ p_{\theta}(y^{i}) \ = DF(w(x|y^{(i)},\emptyset), Z(x|y^{(i)},\emptyset)) + M(\theta,\emptyset;y^{i}) \tag{2}$$

Where the distance between two distributions is determined using the DF divergence function.

$$M(\theta,\emptyset;y^{(i)}) = -DF(w(x|y^{(n)},\emptyset), v(y,\theta)) + R_{w(n|m^{(i)},\emptyset}[log(x^{(n)}|z,\theta)] \tag{3}$$

The KL functions' non-negativity allows equation (2) to be rewritten as:

$$\log p_{\theta}(y^{i}) \geq -DF(w(x|y^{(n)},\emptyset), v(y,\theta)) + R_{w(n|m^{(i)},\emptyset}[log(x^{(1)}|z,\theta)] \tag{4}$$

$$M_{VAE} = \sum_{j=l}^{n}(\tfrac{1}{L}\sum_{m=1}^{m} logp\ (y^{i}w^{(m,n)},\theta) - DF(w(x|y^{(i)},\emptyset), N(0,1) \tag{5}$$

After extracting the features using VAE, the extracted features will be provided to the feature selection model.

3.4. Feature Selection using Correlation Analysis

The best candidate subsets were then selected using a correlation analysis feature selection technique, which reduces and eliminates redundant features. The statistical strength of the relationship between the relative movements of two variables is shown by the correlation coefficient. The values range from -1.0 to 1.0. Correlation between the target and good variables is one of the reasons for using correlation to select features. Furthermore, variables should be correlated with the goal but not with one another. Two variables can be predicted from one another if they are related. Consequently, if two features are associated, the approach only actually requires one of them because the other one does not deliver any additional data. The Pearson Correlation will be applied in the present scenario. The coefficient, ranging from -1 to 1, was computed using the following formula:

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x,\sigma_y} \tag{6}$$

3.5. GMM-Based Clustering

The notation $g \in \{1,\ldots,G\}$, where G is the number of clusters in our dataset, represents a Gaussian mixture, which is a function made up of several Gaussians. Each Gaussian g in the mixture consists of the following parameters: an average μ that specifies its center. The width of the covariance is comparable to the dimensions of a defined ellipsoid in the context of multiple variables. The Gaussian function's shape is determined by the chance of mixing. The GMM algorithm has been selected as the suggested method for clustering the time series data. The remarkable effectiveness of this clustering method in creating clustering rings around the moving nodes is its defining feature.

3.5.1. Determination of the Number of Components in Variables

The number of components in GMMs equals one per variable. GMM is represented as follows:

$$f(a:\theta) = \sum_{x=1}^{h} f_x(a_y;\mu_x;\sigma_x) \tag{7}$$

Where $f(a:\theta)$ indicates the density function of univariate Gaussian mixed distributions, h denotes Gaussian mixture components, π_x denotes mixing weights, and $f_x(a_y;\mu_x;\sigma_x)$ signifies the component probability density function.

3.5.2. Selection of Covariance Type and Initialization Strategy

Estimated parameters are used to compute the likelihood value. Each variable's component counts are established based on the information criterion. The number of observations in each component indirectly affects the covariance matrices and mixing weights in the mixture model. The following is the covariance matrix structure for multivariate GMMs that corresponds to the GMM component clusters in the grid structure:

$$\Sigma_x = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,q}\sigma_1\sigma_q \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{q,1}\sigma_q\sigma_1 & \rho_{m,2}\sigma_q\sigma_2 & \cdots & \sigma_q^2 \end{bmatrix} \tag{8}$$

For $x = 1,\ldots,h$, where q is the number of dimensions of the data collection, and h is the number of components. The p-dimensional standard deviations are denoted by $\sigma_1,\ldots,\sigma_q$. The variables' component correlations are described as $\rho_{1,2} =$

$Corr\ (X1, X2), \ldots, \rho_{q,q-1} = Corr(X_q, X_{q-1})$. Because variables have components of varying sizes, this kind of covariance matrix is utilized. Figure 2 shows the flowchart of the GMM approach.

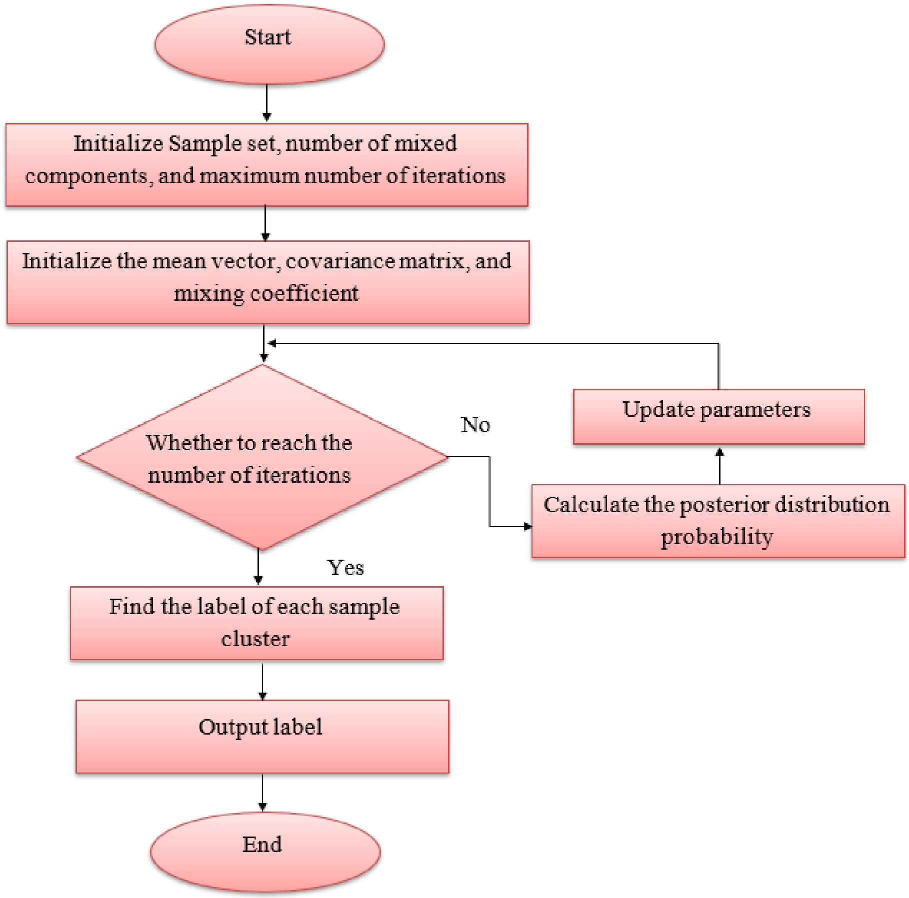


Fig. 2 Flowchart of GMM model

3.5.3. Expectation-Maximization-Based Cluster Formation

A time series is basically an organized collection of observations taken at regular intervals. The process of clustering such data needs a proper decision on multiple aspects, especially the distance metric and the evaluation criteria, since time series data are usually extensive in volume, very dynamic, and naturally high-dimensional. In GMMs, the parameters π, μ, and σ are usually evaluated utilizing the EM algorithm, which is a sequence of the Expectation and Maximization steps; it progressively updates these parameters until convergence occurs. To find the ideal component numbers in mixture models, parameters are estimated using the EM technique.

3.5.4. Load Trend Characterization using Probabilistic Cluster Assignments

Clustering of data allows the discovery of groups/months that share the same EC trends. Furthermore, a more thorough and in-depth comprehension of the behavior of patterns of power consumption can be obtained by load trend characterisation. It makes it possible to identify subgroups with comparable demand patterns, which aids in the development of machine learning-based solutions for a variety of analytical problems involving energy patterns. System planning, load management, maintenance, and marketing all heavily rely on the load trend characterization, which can be performed at various granularities. Seasonal analysis is used to illustrate how patterns of power use change across a year's several seasons and clusters.

Daily analysis shows how power usage habits change over the course of a week. Periodical analysis: To obtain a more thorough trend analysis, the entire daytime length can be split into distinct periods depending on the consumption patterns. Total Consumption Evaluation shows that a state's general pattern of power use over time can serve as a predictor of future demand. Additionally, it will aid in developing strategies for the future and improving resource planning.

3.6. Electric Load Forecasting

The proposed EDFORCE model uses the Adam-optimized A-SCE-LSTM model with the swallow search for forecasting electricity demand.

3.6.1. Attention-Enhanced Serial Cascaded Encoder LSTM Model

In order to model the temporal dependencies of various scales and the nonlinear dynamics in electricity consumption, the A-SCE-LSTM model is utilized in this work. In this structure, multiple LSTM-based encoder modules are stacked up, and each encoder works on the latent representation passed from the preceding one. To boost the retention of information and avoid the loss of features through cascaded propagation, an attention mechanism is interposed between the consecutive stages of the encoder, which helps the network to focus on the most crucial temporal features, like unexpected demand increases or transitional patterns, before the deeper feature abstraction. This serial cascading technique allows the system to learn the hierarchy of time, and consequently, the system is able to extract more of the short-term variations, medium-term cyclic trends, and long-term seasonal dependencies gradually. Figure 3 depicts the architecture of the A-SCE-LSTM model.

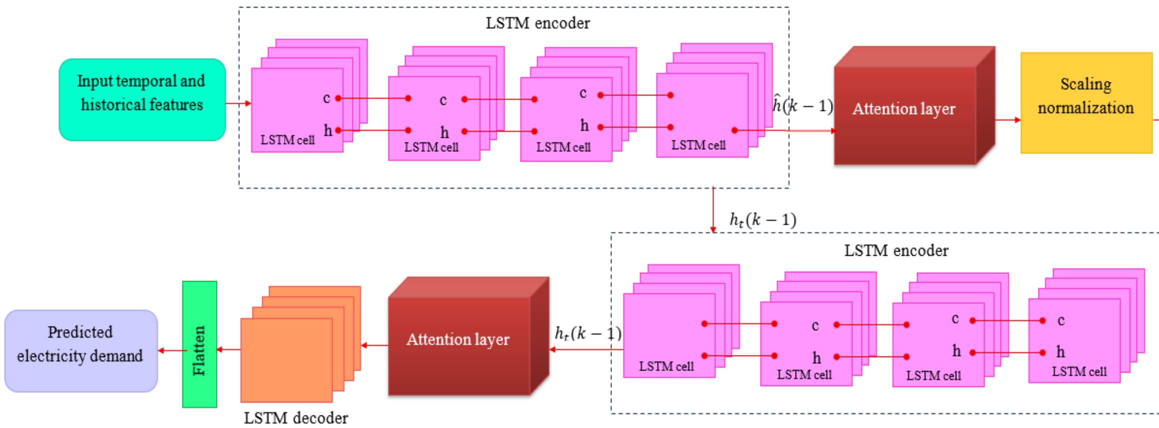


Fig. 3 Attention-enhanced serial cascaded encoder LSTM model

Let the input time series be denoted as

$$\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{T}\} \tag{9}$$

Where x_t is the electricity demand at time t. The first encoder's output is determined as:

$$h_t^{(1)}, c_t^{(1)} = LSTM^{(1)}\left(x_t, h_{t-1}^{(1)}, c_{t-1}^{(1)}\right) \tag{10}$$

Then the hidden state obtained is given to the next encoder stage:

$$h_t^{(k)}, c_t^{(k)} = LSTM^{(k)}(h_t^{(k-1)}, h_{t-1}^{(k)}, c_{t-1}^{(k)}) \tag{11}$$

Where k represents the index of the cascaded encoder, $h_t^{(k)}$ and $c_t^{(k)}$ represent the hidden and cell state, respectively. For enhancing the learning context, an attention-based representation can be expressed as $\tilde{h}_t^{(k-1)}$ is computed as:

$$\tilde{h}_t^{(k-1)} = \alpha_t \odot h_t^{(k-1)}, \alpha_t = \mathrm{softmax}\left(W_a h_t^{(k-1)} + b_a\right) \tag{12}$$

where α_t is the attention weight vector and $\odot$ is the symbol for element-wise multiplication. Then this attention-refined feature is passed to the next encoder stage, which allows better feature prioritization across temporal dimensions. The internal LSTM activities in every encoder layer operate according to the gated structure:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{13}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{14}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{15}$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{16}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{17}$$

$$h_t = o_t \odot \tanh(c_t) \tag{18}$$

Where, i_t, f_t, o_t, c_t, and h_t represents the input gate, forget gate, output gate, updated cell memory state, and updated hidden state. Also, W, and U represent the weight matrices for the input and recurrent connections, and b represents the bias vectors. Finally, the latent representation from the last encoder stage $h_t^{(k)}$ is sent to the decoder network for demand forecasting:

$$\hat{y}_{t+\triangle} = Decoder(h_t^{(k)}) \tag{19}$$

Where Δ is the forecasting horizon, and $\hat{y}_{t+\triangle}$ This is the time when the predicted energy demand is expected to occur.

The encoder layer yields a polished temporal representation, which is an indicator for the forecasting process. The model hyperparameters are then fine-tuned by the hybrid Swallow Search Optimization-Adam mechanism that performs global hyperparameter exploration followed by local gradient-based refinement, hence achieving faster convergence and better prediction accuracy.

3.6.2. Hyperparameter Tuning Using Swallow Search–Adam Optimization

The deep learning model employed in this work has various hyperparameters, including the number of input neurons, hidden layers, window_size, trade-off between flatness and tolerance, stateful, memory_between_batches, output neurons, batch_size, and width ∊. These characteristics must be correctly defined in order to create reliable forecasting models.

The hybrid Swallow Search–Adam Optimization (SWSO–Adam) has been used for hyperparameter tuning, which combines the social communication, migratory, and feeding habits of the swallow search algorithm and the refinement strategy of the Adam optimizer. Figure 4 shows the flow chart of the hybrid SWO-Adam optimization algorithm. The parameters are chosen based on the dimensions of the incoming data and the predicted time horizon. Iterative training and validation runs are used to determine additional parameters, including C, η, number of epochs, and batch_size.

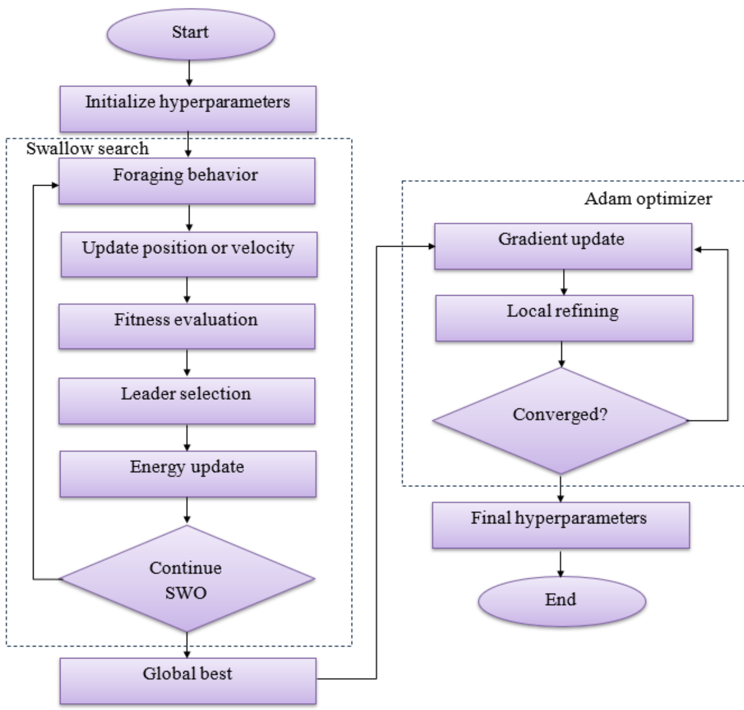


Fig. 4 Flow chart of the SWO-adam optimization algorithm

The algorithm simulates a population of M swallows, each represented by a position vector. To improve exploration, swallows' foraging habits are modeled in the first phase.

In groups, swallows hunt for food, shifting their locations in response to successful foraging experiences. The position of every swallow is set uniformly over the search space:

$$p_{x,y}^{(0)} \sim a(l_y, u_y) \tag{20}$$

where the lower and upper boundaries for dimension y are denoted by l_y and u_y, respectively. The framework makes use of velocity-based updates that are based on momentum-driven exploration instead of completely random movements across the search space. The algorithm also considers past positions in order to improve the agents' navigation and adaptation within the solution space, thereby facilitating their movement more accurately. This promotes a balanced search behavior that improves search efficiency and solution quality by allowing for concentrated exploitation in later stages and widespread exploration in early iterations.

$$V_x^{(s+1)} = \omega . V_x^{(s)} + co_1 . f_1 . \left(q_{best,x}^{(s)} - a_x^{(s)}\right) + co_2 . f_2 . \left(a_M^{(s)} - a_x^{(s)}\right) \tag{21}$$

In this case, the momentum is controlled by $\omega . V_x^{(s)}$, which keeps some of the previous velocity. The cognitive term, $co_1 . f_1 . \left(q_{best,x}^{(s)} - a_x^{(s)}\right)$, signifies the effect of the present leader, motivating the agent to approach the topperforming member of the team. The cognitive learning coefficient is represented by co_1, and f_1 is a random number in [0, 1]. The social term, $co_2 . f_2 . (a_M^{(s)} - a_x^{(s)})$, represents the leader's power, motivating the agent to approach the most productive group member. Subsequently, the agent changes its position according to the newly calculated velocity, thus gradually advancing in the direction pointed out by these collective influences.

$$a_x^{(s+1)} = a_x^{(s)} + V_x^{(s)} \quad (22)$$

By lowering co_1, which decreases exploration, and raises co_2, which improves exploitation of the search space, the framework exhibits optimal dynamic behavior over time and causes a change from independent search toward swarm convergence. The entire flock's behavior pattern alters when a bird shifts from being a follower to a leader or vice versa. Other birds will adjust their places in response to the leader's active guidance of flock movements. The SWSO algorithm encourages high-performing and motivated people by selecting a leader based on a combination of fitness and energy:

$$Lead = arg \min_x (fit_x^{(s)} + (1 - En_x^{(s)})) \quad (23)$$

The energy of each agent is adjusted to mimic exhaustion and partial recuperation.

$$En_x^{(s+1)} = min\left(max\left(En_x^{(s)} - 0.01 + 0.05 . \mu_x, 0\right), 1\right) \quad (24)$$

After a set number of SWO rounds, the global best position $\Theta^{(0)} = \mathrm{G}^{\mathrm{best}}$ is chosen as the next stage in hyperparameter setting. Adam optimizer is used as a local refinement tool further to crack the model parameters in the neighborhood of $\Theta^{(0)}$. Θ is the vector of hyperparameters and trainable parameters (or a subset regarded as most sensitive), and $\mathcal{L}(\Theta)$ is the validation loss (e.g., MAPE or rRMSE) of the EDFORCE model. During the k-th iteration, Adam will update the gradients of the first and second moments of the gradient $\nabla_\Theta \mathcal{L}(\Theta)$:

$$\begin{aligned} \mathrm{m}^{(k)} &= \beta_1 \mathrm{m}^{(k-1)} + (1-\beta_1) \nabla_\Theta \mathcal{L}\left(\Theta^{(k-1)}\right), \\ \mathrm{v}^{(k)} &= \beta_2 \mathrm{v}^{(k-1)} + (1-\beta_2) (\nabla_\Theta \mathcal{L}(\Theta^{(k-1)}))^2, \end{aligned} \quad (25)$$

Where the first and second moments' decay rates are denoted by $\beta_1$and β_2, respectively. The bias-corrected estimates are $\widehat{\mathrm{m}}^{(k)} = \mathrm{m}^{(k)}/(1-\beta_1^k)and\hat{\mathrm{v}}^{(k)} = \mathrm{v}^{(k)}/(1-\beta_2^k)$, and the parameter update rule is as follows:

$$\Theta^{(k)} = \Theta^{(k-1)} - \eta \frac{\widehat{\mathrm{m}}^{(k)}}{\sqrt{\hat{\mathrm{v}}^{(k)}} + \epsilon} \quad (26)$$

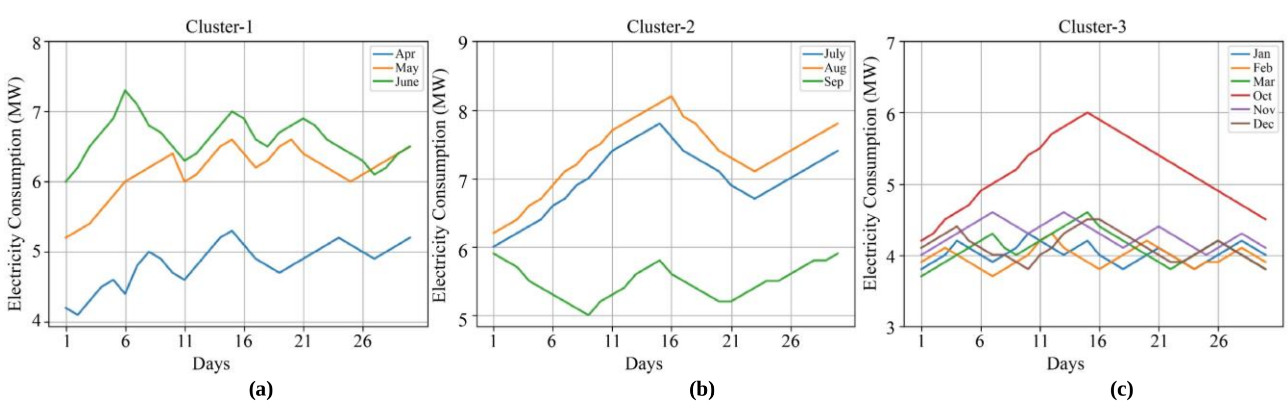
where η is the learning rate, and ϵ is a small constant concatenated for numerical stability. The local refining process that starts from $\Theta^{(0)} = \mathrm{G}^{\mathrm{best}}$ allows the EDFORCE model to find a more exact optimum that SWO alone might not attain because of its population-based, gradient-free nature. The hyperparameter configuration, which was predicted to be optimal through SWO-Adam, is then used to set up the Serial Cascaded Encoder-LSTM model with the advantages of a more stable convergence, quicker training, and better prediction accuracy.

4. Results and Discussion

The suggested EDFORCE framework is implemented with PyCharm Community Edition 2016.3.2 and Python 3.5.2 (64-bit). The trials were performed using a computer with an Intel Core i7-7700 CPU (2.80 GHz), 8 GB of RAM, and an NVIDIA GeForce GTX 1050 graphics card. The EDFORCE model was built on TensorFlow, the open-source deep learning framework developed by Google, and Keras 2.0.8 was used as a primary frontend. The efficiency of the suggested EDFORCE model is tested against old methods, such as K-PCD [24], FNET [29], and VW AA-KELM [30]. The efficacy of the developed EDFORCE approach has been assessed utilizing measures such as AC, PR, RE, F1-s, MBE, MAPE, and rRMSE.

4.1. Performance Analysis

This section analyzes the efficiency of the developed approach in clustering, prediction, and tuning the hyperparameters.



(a) **(b)** **(c)**

Fig. 5 In-Depth clusters analysis

Based on the clustering results, three main groups/clusters can be identified from the whole consumption patterns. The months allocated to each cluster are displayed in Figure 5, which also provides significant detail on the clustering results. It is pretty apparent from the Figure that the three clusters practically coincide with the three main seasonal patterns, including summer, winter, and the rainy season. The actual and EDFORCE-predicted EC values over 30 days are compared in Figure 6. The trend indicates that the EDFORCE model is very accurate in tracking the variations in demand and is thus able to correctly spot rising as well as falling consumption periods. The consistency is a clear indication of the accuracy and reliability of the proposed forecasting framework in detecting nonlinear demand behavior. It is very crucial for operational energy planning and peak-load management.

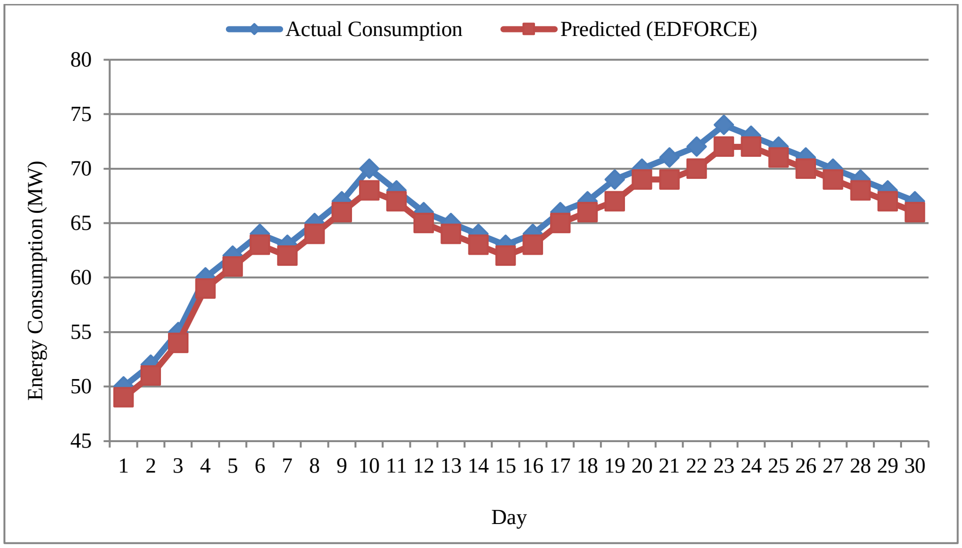


Fig. 6 Actual and predicted energy consumption

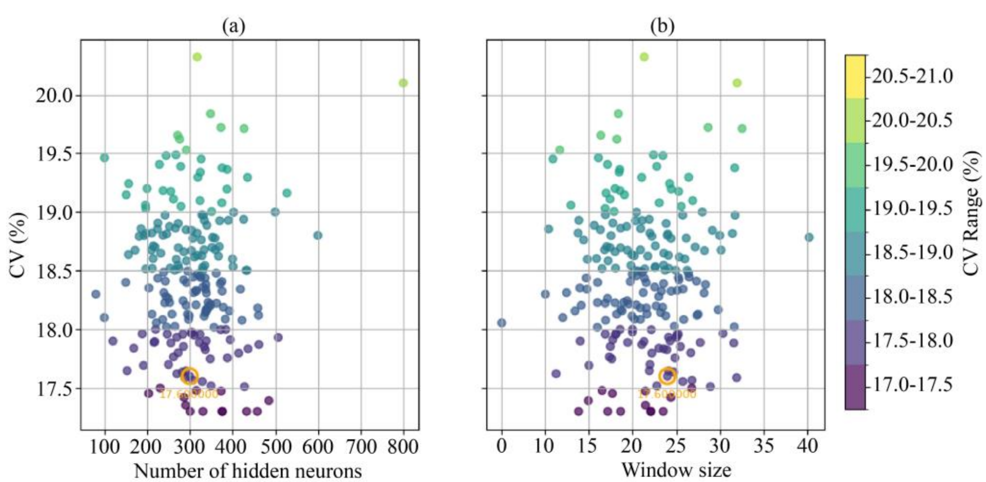


Fig. 7 Scatter plots of the optimization process for the EDFORCE model. (a) Hidden neurons vs. CV (%), and (b) Window size vs. CV (%).

The best individuals from all generations in the SWO-Adam task for energy forecasting are visualized in Figures 7(a) and 7(b). The hidden neurons' count and the window size are compared with the percentage of the Coefficient of Variation (CV).

In Figure 7(a), the SWO-Adam tuned A-SCE-LSTM model is reported to be working at its best with 139 hidden neurons and notching up a prediction accuracy of 17.5%. The optimum window size search is shown in Figure 7(b), and 23 time lags are clarified as being the most informative. It is evident from the figure that the model is strongest when the window size is between 20 and 40 lags and that performance tends to be diminished when values are outside this interval.

4.2. Comparative Analysis

The forecasting outcomes of the developed EDFORCE are compared with previous approaches such as K-PCD [24], FNET [29], and VW AA-KELM [30] in terms of measures such as AC, PR, RE, F1-s, MBE, MAPE, and rRMSE.

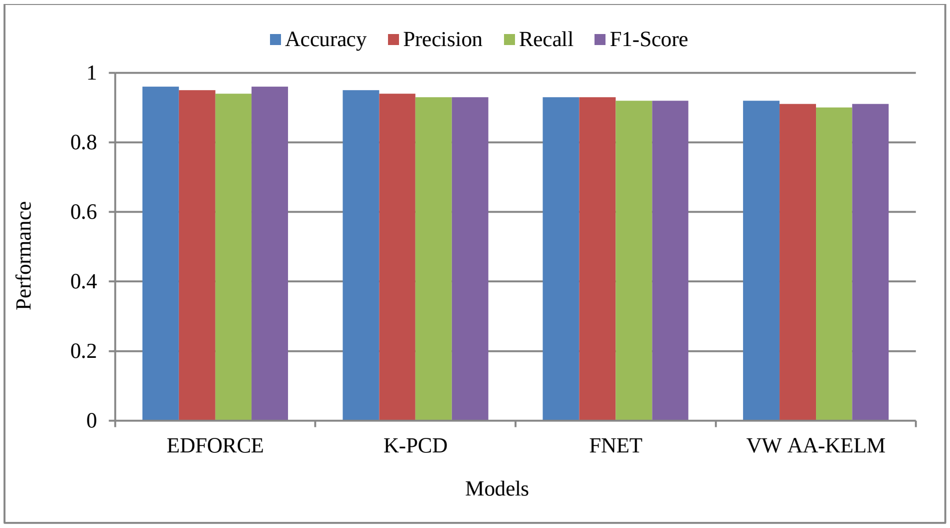


Fig. 8 Performance comparison

A performance comparison for the proposed EDFORCE model against K-PCD, FNET, and VW AA-KELM model with the performance metrics such as AC, PR, RE, and F1-s is presented in Figure 8.

The results show that EDFORCE excels at all the metrics, thus demonstrating the most consistency and robustness in prediction. The F1-score being the highest is an affirmation of the model's good performance on both precision and recall metrics.

In contrast, the accuracy is slightly higher than others, and shows better classification reliability. The proposed EDFORCE model achieves a higher accuracy of 97%, whereas the previous models, such as K-PCD, FNET, and VW AA-KELM model, achieve accuracy of 95%, 93%, and 92% respectively. Figure 9 shows that the MAPE values fluctuate across epochs ranging from 10 to 100 for all models. EDFORCE keeps on showing a lower MAPE throughout the training, which means its forecasting accuracy is superior, along with a smaller relative error. Compared to previous models, the proposed method has advantages due to its rapid convergence and steep decay, indicating that the learning is not only more efficient but also allows efficient parameter tuning.

Figure 10 illustrates the MBE trend throughout the training epochs for all the proposed and the previous models. The proposed EDFORCE achieves the lowest MBE value in all epochs. Thus, the proposed EDFORCE performed excellently regarding bias reduction and the balancing of overestimation and underestimation of energy demand. EDFORCE, as compared to K-PCD, FNET, and VW AA-KELM, showed the fastest elimination of bias levels. This was a clear indication of the success of the serial cascaded learning and optimization approach.

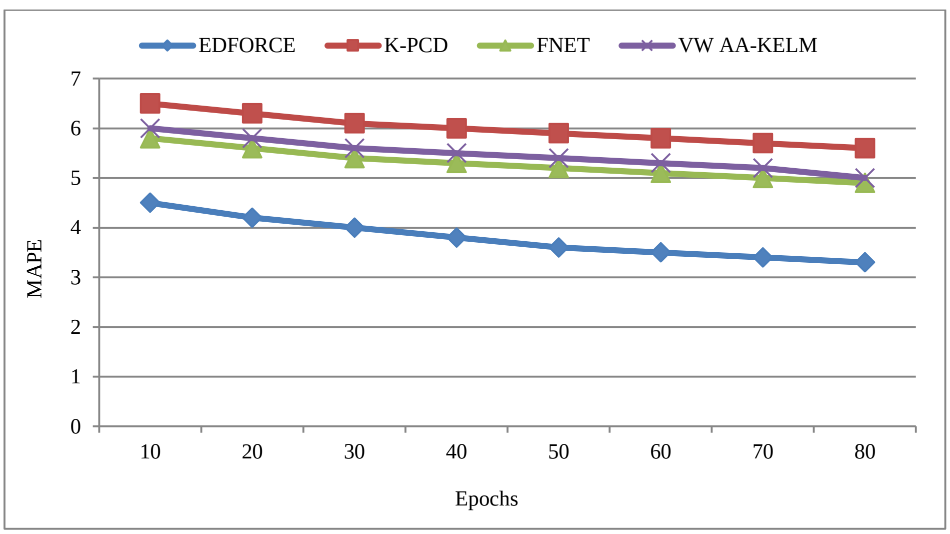


Fig. 9 MAPE comparison

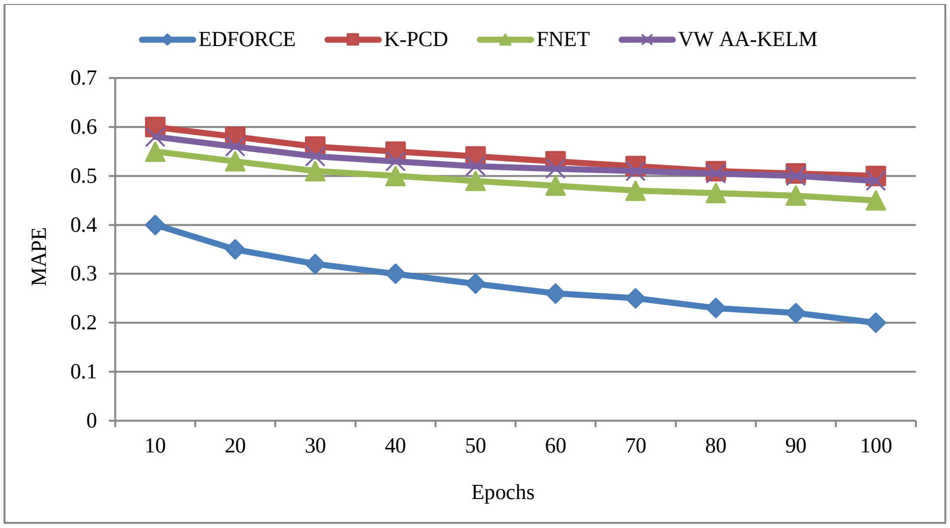


Fig. 10 MBE comparison

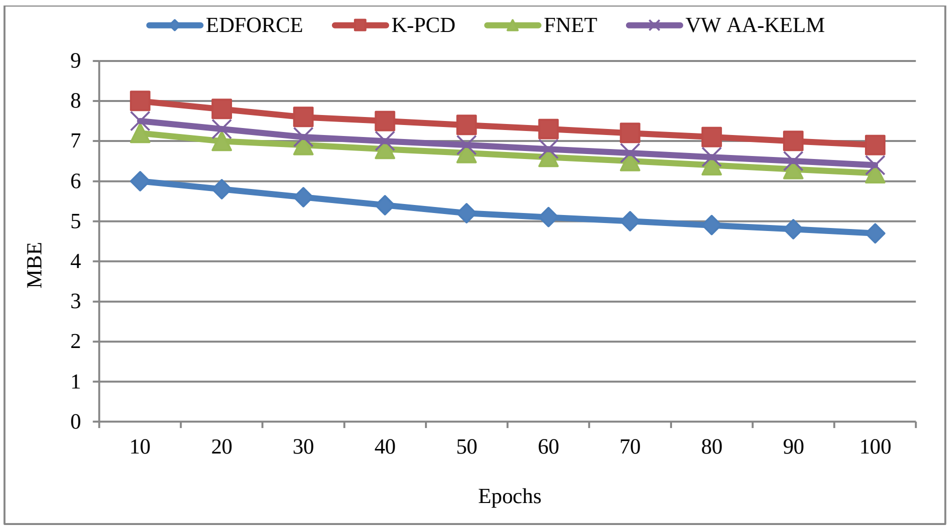


Fig. 11 rRMSE comparison

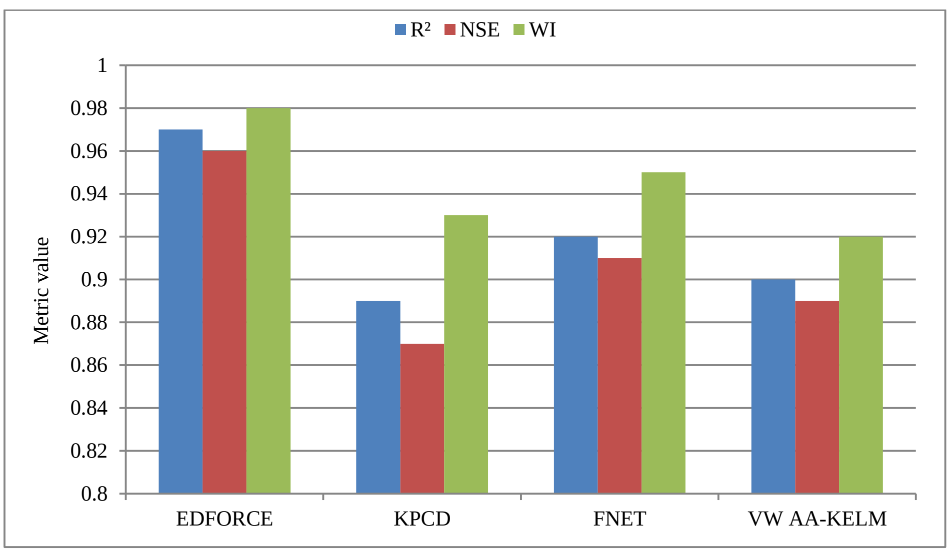


Fig. 12 R^2, NSE, and WI comparison across models

Figure 11 displays the rRMSE values comparison among the different models throughout the epochs. The rRMSE of the EDFORCE model showed a constantly decreasing trend, and it was the best among all models in the whole training period. This implies that the model is very good at prediction deviation control, especially when there are changes in the energy demand conditions. The steadily converged pattern of EDFORCE is an indication of strong generalization ability, while the relatively high rRMSE in the existing methods suggests that they treat the nonlinear temporal dependencies of energy consumption data poorly.

Figure 12 shows the comparison of the developed model with the previous approaches through statistical evaluation measures R^2, Nash–Sutcliffe Efficiency (NSE), and Willmott's Agreement Index (WI). Among the three metrics, the proposed EDFORCE has the highest scores, thus confirming the strongest correlation with the observed demand, the best predictive consistency, and also the closest correspondence with the distribution of the real data. The high WI and NSE values indicate outstanding control over error distribution, whereas the highest R^2 value proves the model's ability to capture variance well.

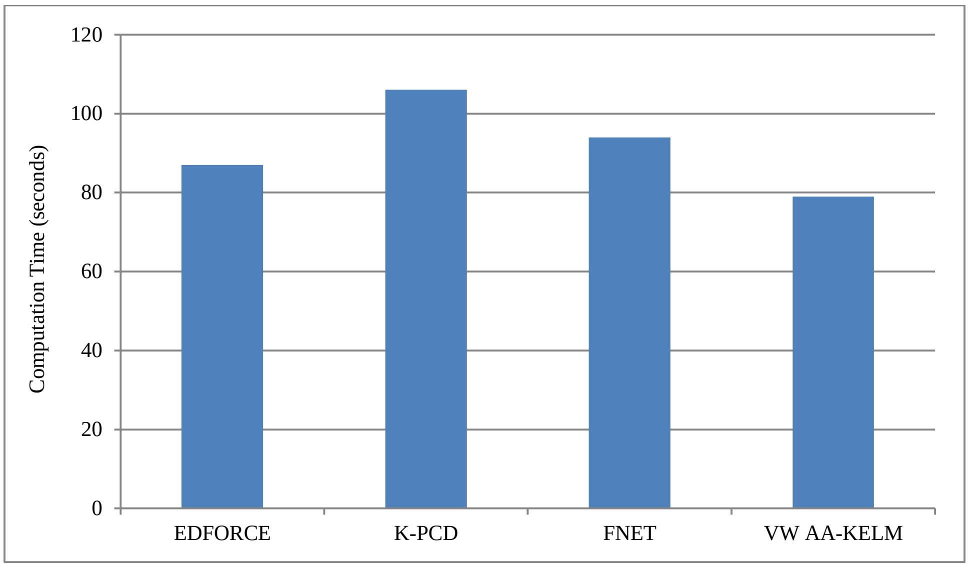


Fig. 13 Computation time comparison

Figure 13 shows the total computation time that each model needs during the forecasting process. The EDFORCE model does need much less computation time compared to K-PCD, FNET, and VW AA-KELM. This is a clear indication of the computational efficiency of the proposed method, which makes it particularly effective in real-time and resource-constrained situations. The proposed EDFORCE model consumed 23.8%, 14.9%, and 1.3% less energy than the K-PCD, FNET, and VW AA-KELM models, respectively.

4.3. Discussion

The EDFORCE framework outperforms K-PCD, FNET, and VW AA-KELM across accuracy and error metrics in the forecast. The main reason for this enhancement is the season-aware GMM segmentation, which leads to a more homogeneous data set, and also the power of the cascading encoder-attention structure that is able to capture both short- and long-term temporal dependencies much better than the traditional single-stage LSTM-based models. To begin with, EDFORCE, with its Gaussian Mixture Model-based probabilistic seasonal clustering partitions the historical demand data into homogeneous seasonal groups. This prevents the mixing of the seasonal pattern, and thus, more stable temporal representations are provided to the forecasting model, which is not the case with the conventional methods since they are trained over mixed seasonal behavior. Second, the Attention-Enhanced Serial Cascaded Encoder-LSTM architecture permits EDFORCE to better tackle the multi-scale temporal dependencies. The cascaded encoder pathway sequentially reduces the load characteristics in short, medium, and long terms. At the same time, the attention mechanism brings the time steps with the most information, such as demand spikes or transitional behavior.

This leads to a stronger nonlinear temporal modelling capability guaranteed over single-stage LSTM or CNN-RNN hybrids. Thirdly, the SWO-Adam hybrid optimization mechanism adds to the global solution search along with local adaptive refinement, which amalgamates and thereby improves training stability and convergence speed.

This, in turn, means that the model parameters are better-tuned and the risk of over-fitting is reduced. The combined improvements led to the optimal performance of the proposed EDFORCE model. The proposed model achieved the most fantastic accuracy, the most minor forecasting error, and the least computational burden as compared to the previous methods.

5. Conclusion

In this paper, a novel Energy Demand Forecasting using Optimized Resilient serial Cascaded Encoder-LSTM (EDFORCE) technique has been proposed. The Gaussian Mixture Modelling efficiently segments the season-based data, which enhances the efficiency for demand forecasting. The novel A-SCE-LSTM model, which is optimized by using the Swallow Search-Adam Optimization algorithm, efficiently predicted the demand with high accuracy. The efficiency of the developed EDFORCE approach had been assessed using AC, PR, RE, F1-s, MBE, MAPE, rRMSE, energy consumption, and computation time. The developed EDFORCE approach attains a higher AC of 97%, whereas the previous models, such as K-PCD, FNET, and VW AA-KELM model, achieve accuracy of 95%, 93%, and 92%. Additionally, the proposed EDFORCE model consumed 23.8%, 14.9%, and 1.3% less energy than the K-PCD, FNET, and VW AA-KELM models, respectively. Future research considers the unification of real-time adaptive learning and edge computing deployment of the EDFORCE model to boost the scalability and thereby increase the practical use in extensive bright grid surroundings.

Acknowledgments

Thanks to all participants and contributors for making this study possible.

References

[1] Ritwik Mohan, and Nikhil Pachauri, "An Ensemble Model for the Energy Consumption Prediction of Residential Buildings," *Energy*, vol. 314, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[2] Bahrudin Hrnjica, and Ali Danandeh Mehr, *Energy Demand Forecasting using Deep Learning*, Smart Cities Performability, Cognition & Security, pp. 71-104, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Razak Olu-Ajayi et al., "Building Energy Consumption Prediction for Residential Buildings using Deep Learning and Other Machine Learning Techniques," *Journal of Building Engineering*, vol. 45, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Taher M. Ghazal et al., "Energy Demand Forecasting using Fused Machine Learning Approaches," *Intelligent Automation & Soft Computing*, vol. 31, no. 1, pp. 539-553, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Nikolaos G. Paterakis et al., "Deep Learning Versus Traditional Machine Learning Methods for Aggregated Energy Demand Prediction," *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Turin, Italy, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jason Runge, and Radu Zmeureanu, "A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings," *Energies*, vol. 14, no. 3, pp. 1-26, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Eunjeong Choi, Soohwan Cho, and Dong Keun Kim, "Power Demand Forecasting using Long Short-Term Memory (LSTM) Deep-Learning Model for Monitoring Energy Sustainability," *Sustainability*, vol. 12, no. 3, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] R. Mathumitha, P. Rathika, and K. Manimala, "Intelligent Deep Learning Techniques for Energy Consumption Forecasting in Smart Buildings: A Review," *Artificial Intelligence Review*, vol. 57, no. 2, pp. 1-33, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jatin Bedi, and Durga Toshniwal, "Empirical Mode Decomposition based Deep Learning for Electricity Demand Forecasting," *IEEE Access*, vol. 6, pp. 49144-49156, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[10] Adebayo Olusegun Aderibigbe et al., "Enhancing Energy Efficiency with AI: A Review of Machine Learning Models in Electricity Demand Forecasting," *Engineering Science & Technology Journal*, vol. 4, no. 6, pp. 341-356, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] G. Vijendar Reddy et al., "Electricity Consumption Prediction using Machine Learning," *E3S Web of Conferences*, vol. 391, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Paulo Vitor B. Ramos et al., "Residential Energy Consumption Forecasting using Deep Learning Models," *Applied Energy*, vol. 350, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Darius Peteleaza et al., "Electricity Consumption Forecasting for Sustainable Smart Cities using Machine Learning Methods," *Internet of Things*, vol. 27, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Javier Manuel Aguiar-Pérez, and María Ángeles Pérez-Juárez, "An Insight of Deep Learning based Demand Forecasting in Smart Grids," *Sensors*, vol. 23, no. 3, pp. 1-30, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mohammad Mahdi Forootan et al., "Machine Learning and Deep Learning in Energy Systems: A Review," *Sustainability*, vol. 14, no. 8, pp. 1-49, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Jihye Shin et al., "Enhanced Data Processing and Machine Learning Techniques for Energy Consumption Forecasting," *Electronics*, vol. 13, no. 19, pp. 1-27, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Ke Yan et al., "Multi-Step Short-Term Power Consumption Forecasting with a Hybrid Deep Learning Strategy," *Energies*, vol. 11, no. 11, pp. 1-15, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Kamal Chapagain et al., "Short-Term Electricity Demand Forecasting using Deep Neural Networks: An Analysis for Thai Data," *Applied System Innovation*, vol. 6, no. 6, pp. 1-29, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Zeynep Hilal Kilimci et al., "An Improved Demand Forecasting Model using Deep Learning Approach and Proposed Decision Integration Strategy for Supply Chain," *Complexity*, vol. 2019, no. 1, pp. 1-15, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Fouzi Harrou et al., "Energy Consumption Prediction in Water Treatment Plants using Deep Learning with Data Augmentation," *Results in Engineering*, vol. 20, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Zhanyu Feng et al., "Energy Consumption Prediction Strategy for Electric Vehicle based on LSTM-Transformer Framework," *Energy*, vol. 302, pp. 1-35, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Mohd Herwan Sulaiman et al., "Utilizing the Kolmogorov-Arnold Networks for Chiller Energy Consumption Prediction in Commercial Building," *Journal of Building Engineering*, vol. 96, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Farhan Hussain, M. Hasanuzzaman, and Nasrudin Abd Rahim, "Multivariate Machine Learning Algorithms for Energy Demand Forecasting and Load Behavior Analysis," *Energy Conversion and Management: X*, vol. 26, pp. 1-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[24] Hao Yang et al., "K-PCD: A New Clustering Algorithm for Building Energy Consumption Time Series Analysis and Predicting Model Accuracy Improvement," *Applied Energy*, vol. 377, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Aditya Pai H et al., "Enhanced Household Energy Consumption Forecasting using Multivariate Long Short-Term Memory (LSTM) Networks with Weather Data Integration," *Results in Engineering*, vol. 27, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Adel Binbusayyis, and Mohemmed Sha, "Energy Consumption Prediction using Modified Deep CNN-BiLSTM with Attention Mechanism," *Heliyon*, vol. 11, no. 1, pp. 1-24, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Shuchen Zhao et al., "Short and Long-Term Renewable Electricity Demand Forecasting based on CNN-Bi-GRU Model," *ICCK Transactions on Emerging Topics in Artificial Intelligence*, vol. 2, no. 1, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Marzia Ahmed et al., "Daily Allocation of Energy Consumption Forecasting of a Power Distribution Company using Optimized Least Squares Support Vector Machine," *Results in Control and Optimization*, vol. 18, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Sujan Ghimire et al., "Explainable Deeply-Fused Nets Electricity Demand Prediction Model: Factoring Climate Predictors for Accuracy and Deeper Insights with Probabilistic Confidence Interval and Point-based Forecasts," *Applied Energy*, vol. 378, pp. 1-34, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[30] Yuqing Wang, and Xiao Yang, "Cloud Computing Energy Consumption Prediction based on Kernel Extreme Learning Machine Algorithm Improved by Vector Weighted Average Algorithm," *2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)*, Xi'an, China, pp. 1-7, 2025. [Google Scholar] [Publisher Link]