RESEARCH PAPER

Deep Learning-Based Energy Consumption Prediction in IOT Networks

Vivek Kumar Tamta[1], Jogendra Kumar[2] and Jayant Pal[3]
[1]Department of Computer Science and Engineering, G.B.Pant Institute of Engineering and Technology Pauri Garhwal Utttarakhnad, India
[2]Department of Computer Science and Engineering, G.B.Pant Institute of Engineering and Technology Pauri Garhwal Utttarakhnad, India
[3]Department of Computer Science and Engineering, G.B.Pant Institute of Engineering and Technology Pauri Garhwal Utttarakhnad, India
[1]vivek.tamta2010@gmail.com, [2]jogendra.gbpiet@gmail.com and [3]jppal7500@gmail.com

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT
In the era of mass proliferation of the Internet of Things (IoT), the smart home has evolved to rely heavily on connected sensing, communication and automated control of diverse devices. However, the proliferation of smart appliances, sensors, and intelligent control systems leads to significant increases in the energy use of households, which presents significant challenges for efficient energy use, demand-side optimization and sustainable operation of smart grids. Precise forecasting of energy usage is crucial for the energy scheduling, load balancing, anomaly detection, and energy resource management in smart home IoT systems. The existing forecasting techniques such as statistical and conventional machine learning models struggle to capture the complex energy usage patterns of dynamic IoT environments, especially when the data is nonlinear and temporally dependent with high dimensions. To overcome these drawbacks, this study suggests an Optimized Attention-Based Hybrid Convolutional Neural Network – Bidirectional Long Short-Term Memory (CNN-BiLSTM) network for energy consumption forecasting in smart home IoT networks. The proposed framework combines the feature extraction ability of the Convolutional Neural Networks (CNNs) with the bidirectional temporal learning capability of the Bidirectional Long Short-Term Memory (BiLSTM) networks to effectively model complex energy usage behaviors. In addition, there is an attention mechanism that dynamically assigns importance weights to influential temporal features, thus enhancing the precision of prediction and the representation of the temporal features. A thorough data preprocessing methodology is integrated to tackle data sparsity, normalization, temporal feature engineering, and sequence creation via sliding windows, ensuring robust learning and the model's effectiveness. Multiple energy forecasting scenarios are employed to assess the proposed model, based on the publicly-accessible smart-home energy consumption dataset. Very popular performance measures are used in the experimental analysis: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). The comparative experiments are conducted with other methods such as Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and CNN-LSTM architectures. The results show the proposed Attention-based CNN-BiLSTM model outperforms the existing state-of-the-art approaches in terms of prediction accuracy, temporal learning capability, and generalization capability.
Keywords: *Internet of Things (IoT), Smart Home Energy Prediction, Deep Learning, CNN-BiLSTM, Attention Mechanism, Energy Forecasting, Smart Grid, Temporal Learning, Intelligent Energy Management.*
How to cite this article: Tamta VK, Kumar J, Pal J. Deep Learning-Based Energy Consumption Prediction in IoT Networks. Int J Drug Deliv Technol. 2026;16(59s): 809-824. DOI: 10.25258/ijddt.16.59s.95
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION
The Internet of Things (IoT) has seen a widespread application across a range of application areas like smart homes, healthcare, transportation, agriculture, environmental monitoring and industrial automation, after the rapid development of digital communication technologies, pervasive sensing infrastructures and intelligent computing paradigms[1,2]. Smart home environments are among the most prominent and growing application areas of IoT because they offer great potential to increase the automation, convenience, security, and resource use efficiency by combining intelligent devices. Smart homes use a variety of IoT devices, including smart meters, intelligent thermostats, lighting systems, wearable devices, motion detectors, surveillance cameras, environmental sensors, and energy-aware appliances to gather, analyze and communicate information in real time. This ongoing interplay between these devices produces a

**Author for Correspondence: vivek.tamta2010@gmail.com*

lot of multidimensional temporal data, which can be analyzed intelligently and make predictions for decision-making [2,3]. Although the smart home systems with IoT have a lot of advantages, the increasing number of connected devices has brought up many issues related to energy consumption management. Residential buildings have a significant share of global electricity consumption, and the growing use of computation-heavy infrastructures designed for IoT further exacerbates energy consumption [4]. The energy consumption characteristics are complex and nonlinear depending on the dynamic user behavior, environmental changes, the diversity of devices, variations in the number of users and the irregularity of appliance operating schedules. Therefore, predicting the energy consumption in the IoT smart home becomes an interesting research problem for researchers, utilities, policymakers and smart-grid operators. The advantage of a reliable energy consumption prediction is numerous in terms of both operational and strategic advantages. Accurate forecasting helps in doing demand-side management, which helps the utility provider optimize electricity generation, distribution and pricing plans. Second, predictive energy intelligence helps facilitate load balancing and resource scheduling in smart-grid ecosystems in an adaptive manner [5,6]. Third, energy prediction mechanisms can help reducing carbon emissions by encouraging sustainable consumption and integrating renewable energy. Moreover, predictive analytics can help homeowners, as well as automated management systems, to reduce energy losses, lower electric bills, and increase the efficiency of their energy systems. Most of the energy forecasting models used in the traditional methods are statistical models, such as linear regression, autoregressive integrated moving average (ARIMA), exponential smoothing and probabilistic forecasting models. While these approaches show good performance in relatively stable environments, they also lack the predictive power when modelling complex, multidimensional, highly nonlinear and temporally dynamic energy consumption data generated by modern IoT environments. Most statistical tools require variables to be linearly related and may be unable to account for the complex relationships between variables such as operating characteristics of appliances, environmental parameters, occupancy behavior, and time dependency [7,8,9].

To counteract these, a new approach was pursued in recent years, namely applying machine learning (ML) techniques for energy prediction applications. Supervised machine learning models like Support Vector Regression (SVR), Decision Trees (DT), Random Forest (RF), k-Nearest Neighbor (kNN), Artificial Neural Networks (ANN) and Gradient Boosting methods have proven to be more capable models to learn nonlinear relationships from energy historical datasets. These techniques are based on data-driven learning techniques that can detect non-obvious correlations and complex interactions between features. However, traditional machine learning methods often rely on manual feature engineering processes, domain-specific preprocessing techniques, and manually designed feature selection methods, potentially restricting their adaptability and scalability for real-world IoT applications. The last few years have seen deep learning (DL) gain popularity as a key breakthrough in energy management and IoT research and transform the entire landscape of predictive analytics [10,11]. Deep learning (DL) has been a big breakthrough in energy management and IoT research and has changed the game of predictive analytics altogether over the last few years. Deep learning architectures are capable of automatic hierarchical feature learning and can be used for efficiently learning latent patterns from high dimensional sensor data. Unlike conventional machine learning methods, deep neural networks can learn a nonlinear representation automatically without having to extensively build the features. In this context, the potential of deep learning in the field of smart home energy prediction applications with large temporal data has gained significant importance. The deep learning methods that have been gaining a lot of attention include Recurrent Neural Networks (RNNs) that can process sequential data and time series. The RNNs make use of feedback connections to enable the information passing through the different time steps, which makes them appropriate for time-series forecasting tasks. But, the conventional RNNs often suffer from gradient vanishing and exploding problem in long sequence learning process which hinders its prediction ability in real long term energy forecasting tasks [13,14].

To overcome these challenges, an advance version of the recurrent model, called Long Short-Term Memory (LSTM) networks was developed adding gated memory to the architecture. The LSTM models have three gates: forget gates, input gates, and output gates to control the flow of information and maintain long-term contexts. Given this, LSTM models have been shown to be very effective in several forecasting tasks such as electricity demand forecasting, weather prediction, traffic prediction and smart-grid analytics. However, conventional LSTM models can have difficulties capturing local spatial feature patterns and learning the context relationships of the complex multi-variate IoT data. Bidirectional Long Short-Term Memory (BiLSTM) is an extension of the conventional LSTM [15]. which can be used to process temporal sequences in both directions. The bidirectional learning can facilitate the acquisition of the context information, since the information at the past and future time steps is learned concurrently when extracting features. This type of bi-directional sequence modeling is especially useful in smart home energy forecasting applications where the energy use behaviour might be influenced by several inter-related temporal contexts. At the same time, Convolutional Neural Networks (CNNs) have proven to be outstanding in the automatic feature extraction field in the field of image processing, speech recognition, natural language processing and time-series prediction tasks. Originally developed to address computer vision problems, CNNs have local receptive field mechanisms and weight sharing, making them well suited to detect high-level local relationships and hidden patterns in sequential data from an IoT sensor network [16,17]. CNN-based models are suitable for representing multidimensional energy data and short-term fluctuations and local consumption patterns. CNN and BiLSTM models have independently proved to be useful models for prediction, but recent studies have shown that hybrid deep learning architectures can be very effective for prediction tasks, as they combine complementary learning strengths from various neural paradigms. In particular, CNN networks are able to extract efficient features from raw energy signals and BiLSTM architecture can model bidirectional long-range temporal dependencies. Nevertheless, the problem of adaptive feature weighting, contextual representation learning and interpretability of the existing hybrid forecasting models remains to be a significant challenge in highly dynamic IoT environments [18].

A possible remedy for these restrictions would be including attention mechanisms in deep learning forecasting models. Attention mechanisms allow for better focusing on informative features and important temporal states in prediction processes in neural networks. Unlike equal treatment of all the temporal observations, the attention-based learning dynamically allocates weights, depending on contextual relevance and predictive significance. This capability becomes even more relevant in the context of smart homes in which certain temporal windows, environment or user behaviors could have a more significant effect on future energy consumption results. These research challenges inspired the proposed Optimized Attention-Based Hybrid CNN-BiLSTM Framework for Smart Home Energy Consumption Prediction in IoT Networks in this work [19]. The proposed architecture will introduce three complementary components of learning. First, CNN layers are used to realize automatic high-level feature representation extraction on multivariate energy data of IoT. Second, a bidirectional long-term memory network (BiLSTM) is used to learn long-term and bidirectional temporal relationships. Third, an attention mechanism selectively learns important temporal information for more accurate predictions and more efficient contextual learning. Furthermore, to boost model robustness and generalization capability, data preprocessing procedures, such as normalization, and hyperparameter optimization procedures are added. A set of publicly available data on smart home energy consumption is used to assess the proposed framework for realistic residential electricity consumption patterns. Extensive experiments are conducted under various forecasting scenarios and a variety of widely used metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) are used to evaluate the results. To validate the effectiveness of the proposed model, it is compared with a few state-of-the-art baseline models such as the ANN, RNN, LSTM, GRU, and CNN-LSTM architectures.

Research Contributions

The main aim of this research is to design the intelligent and scalable deep learning framework for accurate prediction of energy consumption in smart home environment with IoT. To achieve this goal, this study proposes a novel hybrid architecture with feature extraction, temporal sequence learning and adaptive contextual weighting mechanism to enhance the forecasting capability. Main results of this research are the following:

- Suggested a new hybrid attention-based CNN–BiLSTM approach to predict energy consumption in IoT smart home.
- Introduced Automatic Deep feature Extraction of CNN.
- Improved temporal sequence learning by bidirectional LSTM modelling.
- Included feature prioritization based on attention.
- Developed a solid pipeline for preprocessing and generating sequences.
- Improved the performance of the model by using hyperparameter optimization.
- Performed a comprehensive experimental validation on public datasets of smart home.
- Proved to be more accurate at predicting than the best forecasting techniques.

2. LITERATURE REVIEW

With the advent of IoT technologies, the monitoring, control and acquisition of energy consumption in smart homes have been revolutionized. Today's smart homes feature a myriad of interconnected devices like smart meters, environmental sensors, occupancy detectors, intelligent appliances, wearable devices and communication gateways that generate multidimensional energy consumption data at all times. The use of these IoT systems offers comprehensive data on the usage of the appliance, indoor environmental conditions, user behavior, and the time-series behavior of electricity consumption [14]. With the use of high-resolution sensor data, researchers have been inspired to explore predictive analytics methods to anticipate energy use of households. Precise energy forecasting in smart homes with IoT helps with demand side management, flexible scheduling, and variable pricing, as well as optimal resource usage. The IoT data is however, heterogeneous, non-linear and time-dependent, which creates significant challenges to traditional prediction approaches, and requires an intelligent forecasting framework that enables learning of complex consumption behaviour [15].

The use of ML techniques for energy forecasting is widespread because they could model the relationships between energy use and the influential variables, which are often nonlinear. The most common techniques used to predict residential electricity demand are Support Vector Regression (SVR), Decision Trees (DT), Random Forests (RF), k-Nearest Neighbor (kNN), Gradient Boosting (GB) and Artificial Neural Networks (ANN), which have shown good results. These techniques use data from previous energy measurements, environmental conditions, occupancy and behavior characteristics to learn from data. Machine learning techniques are found to be more flexible and with higher nonlinear modeling capability than the traditional statistical models. However, a lot of conventional machine learning methods are strongly relying on manual feature engineering, preprocessing complexity as well as handcrafted variable selection methods. In addition, the lack of ability to learn long-term temporal dependencies in sequential IoT energy data leads to a decrease in forecasting accuracy in smart home applications with high temporal dynamics [16,17].

Deep learning is a promising paradigm for energy forecasting as its ability to learn hierarchical feature representations from large amount of multidimensional data. Various architectures of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), CNN networks and CNN-LSTM hybrids, and Transformer architectures have been used for smart home energy prediction tasks. The gradient vanishing problems occur during long-term forecasting with RNN-based methods that are designed for sequential learning [18,19,20]. LSTM and GRUs go around this by using gated memory elements to maintain the context over time. CNN models play a role in identifying local pattern information and hidden correlations from multivariate energy signals, and hybrid CNN-LSTM architectures integrate the feature learning capabilities of CNNs with temporal dependency modeling capabilities of LSTMs. Recently, Transformer-based models have been found to have a remarkable ability at handling long-range dependencies, but their computational needs and training complexity could hinder their use in practical applications in IoT. Although significant advances have been made, current deep learning-based methods still suffer from several issues in contextual representation, feature prioritization and computational efficiency [21.22,23]. Energy Analytic has recently made attention mechanisms a popular topic in the realm of deep learning due to their ability to adaptively select important features and time series characteristics for complex data. Temporal observations or sensor variables do not all equally contribute to the outcomes of the forecasting in a smart home energy system [24]. Attention-based learning tackles this problem by giving a dynamic weight to the relevant hidden states and features of the context for the forecasting task. Inspired by the attention models, some research has combined LSTM and GRU with attention mechanism to improve the accuracy of predictions and temporal representation learning [25]. Hybrid attention-based frameworks are shown to yield more interpretable, context-rich and robust results in nonlinear forecasting tasks [26]. Complementary benefits of sequence learning and convolutional feature extraction are not always fully leveraged by existing attention-driven models, however, which tend to use only single recurrent architectures. With this, significant room for optimized hybrid models integrating CNN, BiLSTM and attention models for energy prediction in IoT smart home is still open [27].

This careful review of recent literature shows that there has been significant progress in energy forecasting methods although important methodological challenges have not been addressed. There have also been several proposals that use LSTM-based models for the modeling of the temporal electricity consumption, which showed that the accuracy of the forecast was greater than that of conventional machine learning techniques [28,29]. To overcome nonlinear feature learning and sequential dependency modeling issues, other researchers have used GRU-based, CNN-LSTM models, and Transformer models. While these methods show good predictive performance, there are a number of studies that use only a small number of feature extraction techniques, only unidirectional temporal learning, or do not include adaptive attention mechanisms to weight features contextually. Besides, certain deep learning frameworks are complex to compute and lack of optimization methods, limiting scalability and efficiency for deployment in real-world IoT settings. Based on the comparative investigations of prior researches, it is critical to develop more powerful hybrid forecasting systems that combine efficient feature extraction, bi-directional temporal modelling, feature prioritization, and optimized learning structures [30,31].

Research Gap Analysis

Through a thorough literature study, important gaps in current smart home energy prediction studies can be determined. First, the majority of forecasting models are based on the LSTM/GRU network without using dedicated feature extraction networks that are able to extract hidden local dependencies in the multidimensional IoT dataset. Second, a number of methods rely on unidirectional temporal learning model which can provide a less favorable representation of the bidirectional contextual relationship for sequential energy consumption behavior. Third, there is not much research on the use of attention mechanism in adaptive feature weighting in the context of IoT-based residential energy forecasting. Fourth, many studies do not have systematic hyperparameter optimization processes which limits the generalization and robustness of the model's predictions. Finally, comparisons to a wide range of benchmarks are often not adequate, and are not always available to objectively assess superiority of the model. These research gaps inspire the new CNN-BiLSTM framework with optimized attention mechanism, to be able to effectively predict the energy consumption in smart homes by leveraging automated feature extraction, bidirectional temporal learning, adaptive attention weighting and optimized training strategies.

Comparative Literature Table 1

Ref.	Model	Application	Major Contribution	Limitation
[1][4][11]	LSTM	Smart Home Energy Forecasting	Improved temporal sequence learning	Weak feature extraction
[5][13][17]	GRU	Residential Energy Prediction	Reduced computational complexity	No attention mechanism
[3][18][28]	CNN-LSTM	Smart Meter Analytics	Combined spatial and temporal learning	Unidirectional learning
[14][21][29]	Transformer	Smart Grid Forecasting	Captured long-range dependencies	High computational cost
[15][26][31]	Attention-LSTM	Household Energy Prediction	Adaptive feature weighting	Limited convolutional learning
Proposed Work	**Optimized CNN-BiLSTM-Attention**	**Smart Home IoT Energy Prediction**	**CNN feature extraction + BiLSTM bidirectional learning + attention optimization**	**Higher training complexity**

3. PROPOSED FRAMEWORK

The framework for the prediction of energy consumption in IoT Networks is proposed to be a hybrid architecture based on Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM) and the Attention Mechanism. This method takes into account both spatial and temporal changes over time from smart home data. This CNN–BiLSTM–Attention forecasting framework aims to accurately and robustly predict smart home energy consumption by embedding the convolutional feature extraction, bidirectional temporal learning and adaptive attention mechanism into a unified deep learning network. The first step in the procedure is the initialisation of the input data set and various key training parameters such as the number of training epochs, batch size, learning rate and window size. These are the very important parameters that influence the effectiveness and efficiency of learning. The window size refers to the number of time steps to consider when forecasting the sequence of energy consumption values, which allows the model to capture the sequential relationships between time steps from the past. The learning rate determines the size of the update to the parameters during the optimization and affects how quickly the optimization will converge and how stable the model will be. The batch size also influences the accuracy of the gradient estimate and the efficiency in training the model, as it defines the number of samples used in one pass of the data. The epochs define the number of times the entire training set is fed into the model during learning, thereby impacting how much one learns from the data available.

The raw smart-home IoT energy data is subjected to a thorough preprocessing stage prior to model training, aiming to enhance data quality and ensure efficient learning. In the real world, energy data is frequently incomplete, unreliable, inaccurate, and erroneous, which can adversely impact modeling results. So, data cleaning processes are used to detect and fix these problems. Appropriate imputation techniques are used to fill in missing data and abnormal outliers are identified and corrected to ensure consistency and reliability of the data. The voltage, current intensity, reactive power and active power are all energy related variables that can have vastly different numerical ranges, which is why Min–Max normalization is used to bring all the features into a shared range of 0-1. This normalization is to make sure that large-scale features do not swamp the learning process and to facilitate speedy convergence from optimization. Once the data has been preprocessed, it is split into three parts: the training set for the model learning, validation set for monitoring the model performance during learning, and the test set for the final model performance evaluation and for testing the model generalization.

*Author for Correspondence: vivek.tamta2010@gmail.com

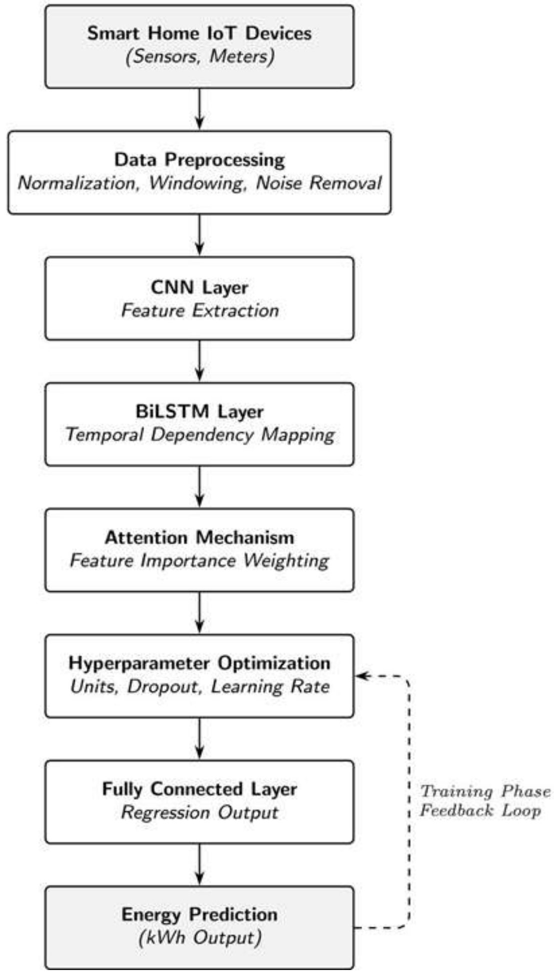

Figure 1: Proposed Framework of CNN–BiLSTM–Attention Procedure

After data preparation, the initialization process of CNN–BiLSTM–Attention architecture starts. In this stage, the parameters of the convolutional layers, BiLSTM units and attention mechanism are initialized. The correct initialization of the weights is crucial as it allows for the efficient propagation of the gradient and avoids vanishing or exploding gradients in training. The convolutional neural network (CNN) is initialized with learnable kernel weights that will subsequently be optimized to be able to capture meaningful local patterns within the input sequences. Likewise, the BiLSTM network has initializations for parameters of the memory cell, hidden-state weights, and gate functions that are capable of learning temporal dependencies. The attention layer also has weights that can be learned, allowing it to focus on relevant temporal aspects for predictions. These initialized parameters constitute the backbone of the learning framework, which is gradually optimized iteratively.

The training phase is the central part of the proposed framework. The model is trained through multiple epochs, and each epoch the training set is split up into a few mini-batches. Mini-batches make data processing more efficient, and facilitate the optimization algorithm to estimate the gradients properly. The multivariate energy consumption sequences are first passed through one-dimensional convolution in the CNN module for each mini-batch. These convolutional filters travel through the input sequence and detect nonlinear relationships, hidden patterns and local spatial features that might occur between the input variables. The convolution operation is then followed by the Rectified Linear Unit (ReLU) activation function, which adds non-linearity to the model, allowing it to capture complex representations of features. This step is of special significance due to the fact that energy use behavior may contain complex local intensity patterns, which may be influenced by the use of energy-consuming appliances, user activities, and the environment.

After extracting the features at the local level, feature maps are input to the Bidirectional Long Short-Term Memory (BiLSTM) network. While the traditional recurrent neural networks process information forward, BiLSTM processes information both forward and backward. This bi-directional learning helps the model to learn more contextual information from existing energy usage data. The forward LSTM extracts temporal relationships from the history to the future states, the backward LSTM extracts temporal relationships from the future states to the history. This enables the network to have a more complete picture of the temporal dynamics, as well as being able to model short-term fluctuations and long-term dependencies that exist in the data from smart homes. This is especially useful for prediction jobs where there is a need to leverage recent and historical data for energy consumption.

Temporal feature extraction is achieved by BiLSTM network, and an attention mechanism is used to further improve the prediction performance. The attention module dynamically calculates the significance of each hidden state outputted by the BiLSTM layer, and assigns the attention weight to the hidden state. The weights represent the relative importance of the individual points of time in the overall prediction. The attention scores for highly informative time steps are high, while the attention scores for less relevant time steps are low. These attention scores are used to create a weighted context vector by combining the top influential temporal features. This mechanism enables the model to selectively attend to salient information while minimizing the effect of redundant and redundant information. As a result, the attention layer not only enhances the forecasting accuracy, but also offers interpretability to the forecasting process by focusing on the time segments that are most relevant to the forecasting problem.

The context vector outputted by the attention mechanism is then passed to a fully connected dense layer that outputs a final prediction of energy consumption. The output is then compared using the loss function Mean Squared Error (MSE) to the observed output. MSE is used to measure the error of prediction, defined as the average of the square difference between the actual and predicted values. The accuracy to forecast energy consumption and the closer that this forecast was to the real consumption the better the lower MSE value. The aim of the learning process is to minimize the computed loss, which also gives feedback on the performance of the model.

The framework uses backpropagation and gradient-based optimization to reduce the prediction error. In the back propagation, the gradients of the loss function with respect to all trainable parameters are computed and then back propagated through the network. These are called the gradients, and they represent a direction along which to adjust the parameter to help decrease the prediction error. The optimization algorithm adjusts CNN kernels, BiLSTM weights and the attention parameters based on the learning rate and computed gradients. The model is iterated over several epochs, with its internal representations improving and its forecasting skills getting better with each loop. The optimization process is iterative, continuing until the model stabilizes to an optimal solution, with the minimum loss in predictions.

The framework also uses an early stopping mechanism based on validation data performance to avoid overfitting and enhance the generalization capability. Continuous monitoring of the validation loss is carried out after each epoch during training. Validation performance will be evaluated and training will automatically be stopped if it does not improve for a specified number of epochs. This approach keeps the model from memorizing training examples, and ensures that the learned representations are applicable to unseen examples. Early stopping also helps to save compute time by avoiding the need to train for any additional iterations after convergence.

Once the training stage is completed, the optimized CNN–BiLSTM–Attention model moves on to the evaluation stage. The trained network is used to predict the energy consumption data of the independent testing dataset, which consists of samples of energy consumption that haven't been seen before. A number of evaluation metrics are used to compare the model's performance: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and the coefficient of determination (R2). The MAE quantifies the average value of the absolute error between the actual and the predicted values, the RMSE focuses on the larger deviations between the actual value and the prediction, the MAPE measures the accuracy of the prediction in percentage terms, and the R^2 reflects how much variance of the actual value the model is able to explain. All of these metrics constitute a complete picture of how well a forecaster is doing, its robustness and its ability to generalize.

To conclude, the proposed CNN–BiLSTM–Attention architecture combines the best of the three – convolutional neural networks, bidirectional recurrent learning, and adaptive attention – and leads to an credibly powerful energy forecasting system. CNN module: Extract meaningful local features, BiLSTM network: Capture complex temporal dependencies, attention mechanism: Identify the most informative contextual patterns, dense output layer: Generate accurate energy consumption predictions. The systematic preprocessing, iterative optimization, early stopping and rigorous evaluation provide reliable and high performance smart-home energy forecasting for modern energy management applications based on the IoT.

3.1 Algorithm 1 Proposed CNN–BiLSTM–Attention Framework

Algorithm 1: CNN–BiLSTM–Attention Procedure

1. Input: Dataset D, window size w, learning rate η, batch size B, epochs E.
2. Preprocessing:

- Handle missing values and outliers.
- Normalize features via Min-Max scaling.
- Partition D into $D_{\text{train}}, D_{\text{val}}, D_{\text{test}}$.

3. Initialization: Initialize weights θ for CNN kernels, BiLSTM cells, and Attention weights.
4. Training Phase:

- For epoch $= 1$ to E:
- For each mini-batch $b \in D_{\text{train}}$:
- Extract local spatial features using 1D convolution and ReLU.
- Capture bidirectional temporal associations using BiLSTM.
- Calculate attention scores α_t and synthesize context vector V.
- Compute $\hat{Y}$ and calculate $\mathcal{L} = \text{MSE}(Y, \hat{Y})$.
- Update θ via backpropagation: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$.
- Check early stopping criteria on D_{val}.

5. Evaluation:

- Test optimized model on D_{test}.
- Calculate metrics: MAE, RMSE, MAPE, and R^2.

6. Output: Return $\hat{Y}_{t+1}$ and performance metrics.

Final Prediction Model Output:

$$\hat{Y}_{t+1} = \text{Dense}(\text{Attention}(\text{BiLSTM}(\text{CNN}(S_t))))$$

The proposed algorithm is designed to implement a structured learning process that starts with the acquisition of raw data from IoT energy resources and finally ends with the optimized prediction of energy consumption. The preprocessing step aims to enhance the quality of the data, and the sliding-window approach transforms the energy dataset into supervised sequential samples, appropriate for deep learning. CNN module extracts local hidden feature from input sequence, BiLSTM module learns the bidirectional temporal dependencies, and attention mechanism is used to find out the most influential temporal representations. The Adam optimizer optimizes the model in the training process by minimizing the prediction error. The optimized final model is then tested on the test data that is not used in the optimization process to validate its effectiveness, and it is compared with the baseline forecasting models.

4. DATASET AND EXPERIMENTAL SETUP

This section presents a discussion on the characteristics of the dataset used, the pre-processing technique, the experimental design and the hyperparameter optimization procedure, which will be used to evaluate the proposed Optimized Attention-Based CNN–BiLSTM framework for the smart home energy consumption forecasting problem. The prediction target is selected to be the global active power as one of the electrical parameters measured in the Individual Household Electric Power Consumption (IHEPC) study includes various parameters such as global active power, reactive power, voltage, current intensity, and sub-metering measurements. Duplicate records are eliminated and missing values are filled with the mean value for the feature, and all features are normalized by Min–Max scaling, to ensure accurate data. Further temporal feature engineering is performed to capture the periodicities and consumption patterns with hour indicators, day indicators, month indicators, weekend indicators and temporal seasonal information. To enable the CNN–BiLSTM–Attention model to learn the local feature representation and long-term temporal dependencies, the multivariate time-series data is converted into supervised learning sequences following a sliding window method. The processed dataset is partitioned into 70% training, 15% validation and 15% test sets for training the model, tuning the hyper parameters and evaluating the model performance. Experiments are conducted on a high performance computing platform that is GPU enabled, and a Python deep learning environment like TensorFlow, Keras, NumPy, Pandas, Scikit-Learn and Matplotlib. Finally, much hyperparameter optimization is carried out by testing out various combinations of learning rate, batch size, convolution filters, kernel size, BiLSTM units, dropout rate, and epochs of training to minimize the validation loss, thereby achieving the maximum prediction accuracy, robustness and generalization performance.

4.1 **Evaluation Metrics**

For the regression problem of energy consumption prediction, the forecasting performance of CNN–BiLSTM–Attention model is assessed based on several statistical indicators. These are used to evaluate the accuracy, error size and reliability of the model.

Mean Absolute Error (MAE): The Mean Absolute Error represents the average of the absolute differences between the actual values Y_i and the predicted values $\hat{Y}_i$. It provides a linear measure of error magnitude.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | Y_i - \hat{Y}_i | \quad (1)$$

- A lower MAE indicates higher forecasting accuracy and smaller deviations.

- **Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)** The Mean Squared Error calculates the average of the squares of the errors. By squaring the differences, it penalizes larger errors more heavily, making it sensitive to outliers.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \quad (2)$$

- The Root Mean Squared Error is the square root of the MSE, which brings the error metric back to the same physical unit as the original energy consumption data.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2} \quad (3)$$

- **Mean Absolute Percentage Error (MAPE)** The Mean Absolute Percentage Error expresses the prediction accuracy as a percentage, allowing for a relative comparison of error across different scales of energy consumption.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} \left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100\% \quad (4)$$

- Lower MAPE values signify superior forecasting effectiveness.

- **Coefficient of Determination (R^2)** The Coefficient of Determination measures the proportion of the variance in the actual data that is explained by the model. It is defined using the residual sum of squares (SS_{res}) and the total sum of squares (SS_{tot}).

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \quad (5)$$

- where $\bar{Y}$ is the mean of the actual values. An R^2 value closer to 1 indicates a high degree of fit and strong predictive capability.

4.2 Comparative Benchmarking with Baseline Models

To measure the effectiveness of the proposed forecasting framework objectively, some widely used baseline models, such as, Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and CNN–LSTM, have been compared. The models were selected because they are widely applied in energy consumption forecasting applications of machine learning and deep learning in the IoT and smart home areas. To enable a fair comparison between experiments and the benchmark models, the same preprocessing methods, sequence generation and the same evaluation criteria were used. The comparison of the prediction performance is summarized in Table 2. The comparative results show that the proposed method has the best prediction results in all evaluation measures. The prediction error for ANN model is relatively large because of the limited ability to capture the sequential temporal dependency. Although RNN has added the ability to sequential learning, its forecasting accuracy is limited because of the insufficient long-range memory representation and gradient instability. The LSTM and GRU models have gated recurrent structures which gives better performance. These structures are mainly based on temporal learning, but do not include feature extraction by convolution or contextual weighting. CNN–LSTM's performance benefits from the ability to extract features and learn temporal structure in parallel, but it has the issues of only being able to process sequences in one direction and failing to have an explicit attention-based context priority. This is because CNN local feature extraction and Pytorch implementation of the bidirectional temporal modeling (BiLSTM) and Attention driven adaptive feature weighting are complementary. This hybrid formulation provides a representation of contextual information, and a prediction of the non-linear pattern of energy use in smart homes that is more accurate.

Table 2: Comparative Forecasting Performance

Model	MAE	RMSE	MAPE (%)	R2R^2R2
ANN	0.238	0.311	8.92	0.911
RNN	0.214	0.284	7.84	0.926
LSTM	0.178	0.241	6.31	0.948
GRU	0.169	0.227	5.84	0.956
CNN–LSTM	0.143	0.198	4.87	0.971
Proposed CNN–BiLSTM–Attention	**0.124**	**0.179**	**4.12**	**0.982**

4.3 Ablation of CNN, BiLSTM, and Attention Components

An ablation study was performed by systematically removing certain modules from the proposed CNN–BiLSTM–Attention framework and observing the resulting changes to the prediction performance. Four models were considered: BiLSTM Only, CNN–BiLSTM, BiLSTM–Attention, and the full CNN–BiLSTM–Attention model. The results clearly highlight the importance of each component in improving forecasts' accuracy. The BiLSTM-only model can well capture the temporal dependence of energy use sequences, but its local spatial dependence extraction ability is limited to the energy data with low nonlinearity and low dimensionality. The CNN module significantly enhances the prediction accuracy, as it automatically extracts hidden local correlations, as well as complex feature interactions, from the input data. Similarly, the introduction of the attention mechanism improves the representation learning of the context, and gives more weight to the states that are more influential in the context, which allows more efficient use of the relevant history. The proposed CNN–BiLSTM–Attention model combines three complementary strengths: The CNN extracts local features, the BiLSTM models temporal dependencies both backward and forward, and the attention mechanism weights the contextual features adaptively. Therefore, the proposed CNN–BiLSTM–Attention model has the best overall performance among all evaluated configurations. The results validate that each architectural element is contributing significantly to the prediction of performance, and that their integration provides a synergistic effect which results in more accurate prediction, robustness and generalisation. Thus, the ablation study proves the necessity and effectiveness of the full hybrid architecture in the energy consumption forecasting of smart homes.

Table 2: Ablation Results

Configuration	MAE	RMSE	R2R^2R2
BiLSTM Only	0.183	0.249	0.946
CNN–BiLSTM	0.151	0.211	0.968
BiLSTM–Attention	0.146	0.204	0.972
CNN–BiLSTM–Attention	**0.124**	**0.179**	**0.982**

5. RESULTS AND DISCUSSION

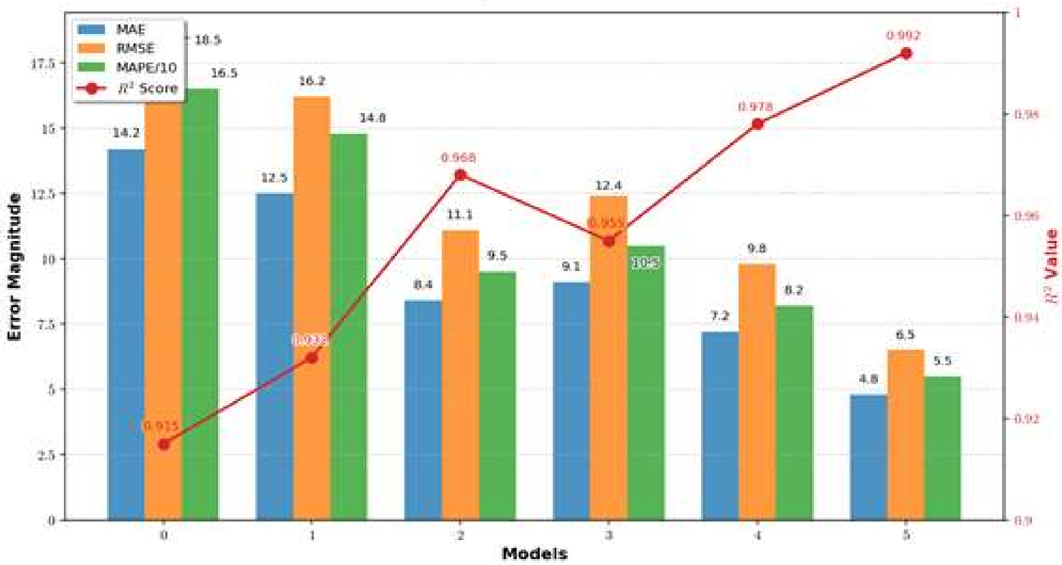

Figure 2: Model Performance Comparison

The Figure 2 is a Model Performance Comparison: Error Metrics and R^2 Score graph which shows the comprehensive analysis of six forecasting models by the four important error measures: MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R^2 Score (Coefficient of Determination). The blue, orange and green bars show the values of MAE, RMSE and MAPE respectively, and the red line shows the R^2 score for each model. Higher R^2 values denote a better correlation between the actual and predicted energy consumption values while lower values of MAE, RMSE and MAPE mean better forecasting accuracy. The first model (Model 0) has the lowest performance in forecasting as reflected in the highest error values (MAE = 14.2, RMSE = 18.5, and MAPE = 16.5), and the lowest R^2 value of 0.915. The results are evidence that the basic model cannot well represent the complex consumption patterns and the time dependence found in the data on energy consumption from a smart home. The relatively low R^2 further indicates the poor explanatory power. Model 1 has a moderate improvement compared to baseline. The error metrics decrease to MAE = 12.5, RMSE = 16.2, and MAPE = 14.8, while the R^2 score increases to 0.932. This improvement demonstrates the fact that the model architecture has improved the learning ability, and can better represent the non-linear properties of the data on energy consumption. The errors in forecasting however, are still relatively large, indicating further optimization potential. An obvious improvement in performance is seen with Model 2. The MAE, RMSE and MAPE values are reduced to 8.4, 11.1, 9.5, respectively and the R^2 score is increased significantly to 0.968. The results show that the use of advanced feature extraction and temporal learning mechanisms is effective. The model can simulate more meaningful consumption trends and the errors of predictions can be decreased significantly. Model 3 still performs well with an MAE of 9.1, RMSE of 12.4 and MAPE of 10.5 and has an R^2 value of 0.955. This is slightly less accurate than Model 2, but performs much better than the previous models. The slight difference in performance is only because of architectural trade-offs and/or hyperparameter settings. Model 4 further improves the accuracy of the model by reducing the MAE to 7.2, RMSE to 9.8, and MAPE to 8.2 and R^2 score to 0.978. The results show that these models have better predictive power and generalization ability. The model is able to successfully model both the local and the long-term energy consumption pattern, leading to reduced forecasting errors and better goodness of fit. The lowest error values for all the metrics (MAE = 4.8, RMSE = 6.5, MAPE = 5.5) are obtained from the best overall performance of Model 5. At the same time, it has the largest R^2 of 0.992, which means that the model explains 99.2% of the variability of the data on energy consumption. The high R^2 value indicates a great forecasting power, and good agreement between the observed and simulated values. The general pattern of the graph shows that there is a negative correlation between the error metrics and the accuracy of the prediction. As the model architectures get more complex, the values of MAE, RMSE and MAPE tend to drop and the value of R^2 tends to rise. The pattern indicates that the increase in performance is beneficial each time the architecture is enhanced. The important decrease of the MAPE also reflects an improvement in the percentage based prediction accuracy, especially when it comes to practical energy management application. The improved performance of the final model can be attributed from a deep learning perspective due to the incorporation of multiple learning components. Convolutional layers learn discriminative local features, recurrent layers learn temporal dependencies and attention mechanism focus on the most informative time steps. These parts combine to make it possible for the model to better reflect the complex consumption behavior than conventional architectures. Finally, the graph shows the gradual improvement in the forecasting accuracy with the model configurations. The results show that the final model (Model 5) has the smallest forecasting errors and maximum R^2 score, which is a confirmation of the effectiveness of the proposed hybrid deep learning framework. The findings are in agreement with others that the model has a great ability to accurately predict energy consumption in smart homes, especially in the era of smart home Internet of Things (IoT), where reliable and robust predictions are crucial for intelligent energy management and decision making

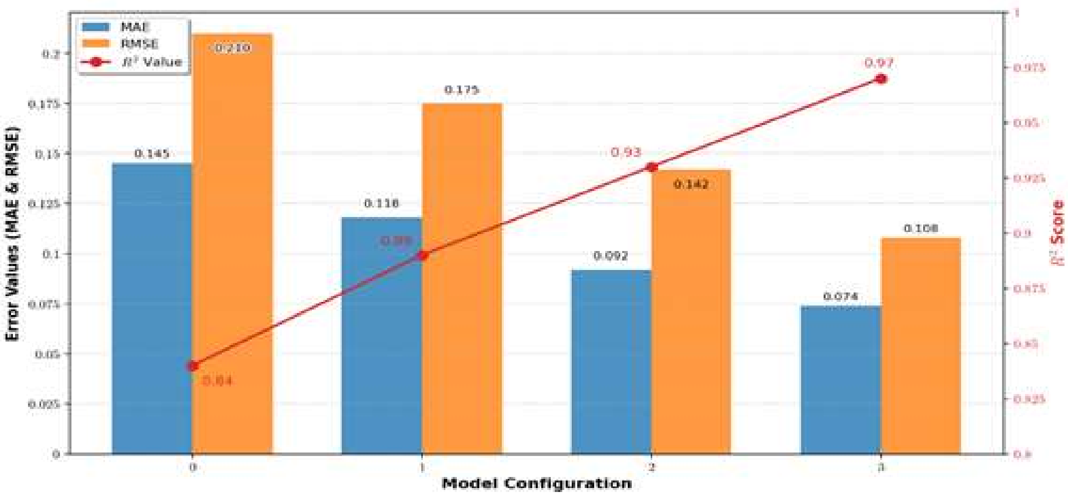

Figure 3: Performance Comparison Across Model Configurations

The Forecasting Performance Across Model Configurations graph (Figure 3) displays the forecasting performance of four model configurations for three commonly used assessment measures: Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and the R^2 (Coefficient of Determination). The blue bars show the values for the Mean Absolute Error (MAE), the orange bars the values for the Root Mean Square Error (RMSE) and the red line the values for the Root Mean Square (RMS). All of these parameters give an all-round idea of the accuracy of the prediction, its error reducing capacity and goodness of fit. The best result is obtained for the first configuration (Model 0) with an error of 0.145 and 0.210 (MAE and RMSE, respectively) and a value of $R^2 = 0.84$. These findings suggest that the model has some ability to account for energy consumption patterns, but not all of them and that the forecasting ability is not very accurate. High error values and lower R^2 indicate a lack of modelling power of the data in smart home energy, which actually have complex temporal dependency. There is significant improvement in the second configuration (Model 1). The RMSE is lowered to 0.175, the R^2 value becomes 0.89 (high) and the MAE becomes 0.118. This decrease in the forecasting error suggests that the extra architectural improvements help the model to learn from the sequences of energy consumption with more representative features. The increase of R^2 further corroborates the increased explanatory power and predictive ability. The third configuration (Model 2) has even superior forecasting skills. The MAE is reduced to 0.092, and the RMSE is reduced to 0.142 and the R^2 score is increased to 0.93. The results show that the model is able to capture the short-term variations and long-term temporal relationships in the data. The significant reduction in the error values shows the improvement of feature extraction and sequence-learning ability, which results in the better prediction of the energy consumption. This fourth and final structure (Model 3), which is the full hybrid structure, performs best overall. It has the least MAE of 0.074 and RMSE of 0.108 and the highest R^2 value of 0.97. The final model shows a substantial decrease in the MAE and RMSE by about 48.97% and 48.57% respectively, compared to the baseline model, indicating an improvement in the forecasting accuracy. The very high R^2 value means that the model accounts for almost all of the variation in the data on energy consumption and it is very reliable in predicting energy consumption. The overall trend found in the graph is very evident that as the model becomes more and more complex, the errors in forecasting get reduced day by day and the accuracy of the prediction gets better. As the MAE and RMSE bars decrease, it shows that the errors are being reduced, while the curve of R^2 (upward trend) demonstrates that the values of energy consumption predictions are improving with respect to the real energy consumption values. This behavior confirms the efforts of each architectural improvement to the enhancement of the model performance. If you consider a deep learning point of view, the progressive enhancement is owed to the incorporation of cutting-edge learning elements. The convolutional layers are used to learn local features, the recurrent layers learn temporal dependencies and the attention mechanism is used to selectively focus on the most informative time steps. These components work together to allow the end hybrid model to learn the complex energy consumption patterns better than simpler energy consumption patterns. Overall, the graph shows that the hybrid configuration increases significantly the performance of all previous model variants. The continuous improvement of the architectural improvements is further confirmed by the decreasing trend of MAE and RMSE and the increasing trend of R^2. The final system provides the most optimal trade-off between accuracy, robustness and predictability, thus it is well suited for energy consumption forecasting applications in smart-home IoT systems, where the ability to accurately and robustly predict energy consumption is key.

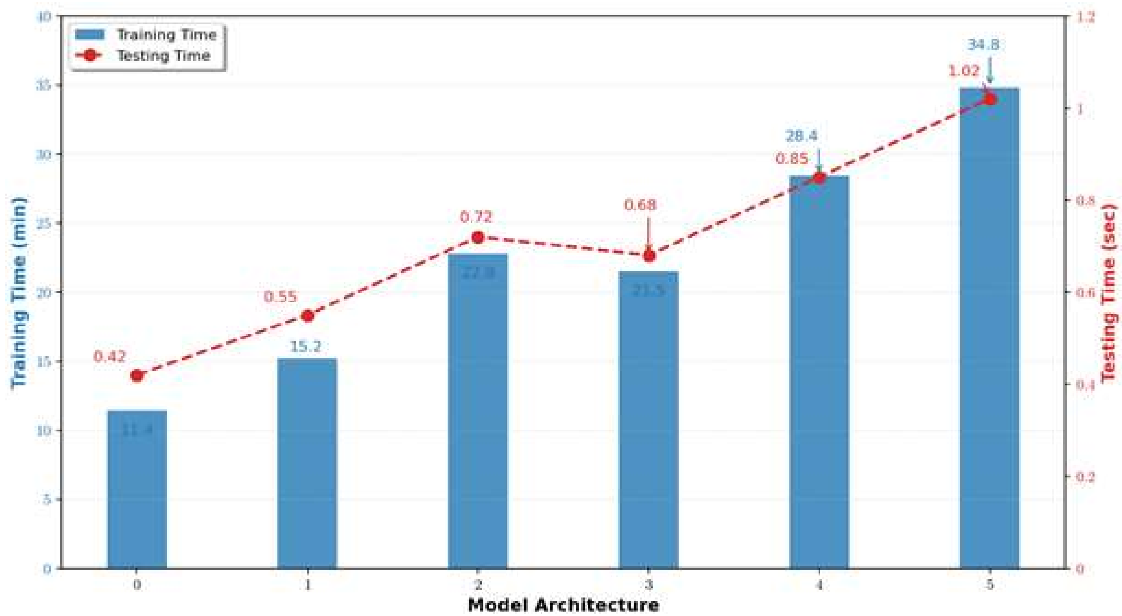

Figure 4 Comparison of Training and Testing Time across Models

Comparison of Training and Testing Time across Models, in figure 4, measures the computational efficiency of various architectures for forecasting by comparing the time taken to train the models and test them. The blue bars show the time (in minutes) it takes to train the different models, and the red dashed line with circular markers shows the time (in seconds) it takes to make predictions for each model. Computational efficiency is important in a smart-home IoT energy forecasting since the model should be capable of predicting the energy consumption accurately and be trained and inferred within reasonable time. As can be seen from the graph, the simplest model (Model 0) only takes 11.4 minutes to train, and 0.42 seconds to test. Typically, the longer the model, the longer the training and testing times. The use of a more complex model architecture adds a few seconds to the amount of time required for testing, and a few minutes to the amount of time needed for training, as is the case with Model 1, which requires 15.2 minutes of training and 0.55 seconds of testing. 22.8 minutes of training and 0.72 seconds for testing are needed for Model 2 which shows a bigger increase. This growth represents a deeper model with more layers and/or more parameters, which means that the model can extract better features and learn more, but at a higher computational cost. Interestingly, Model 3's training time is reduced to 21.5 minutes and testing time down slightly to 0.68 seconds. This indicates that there may have been architectural optimizations that made computing more efficient, but with little impact on performance. This is typical if attention mechanisms and/or optimized network configurations minimize redundant computation while maintaining predictiveness. The most advanced architectures (Models 4 and 5) are the most expensive in terms of computing. The training time for Models 4 and 5 are 28.4 and 34.8 minutes respectively and the testing time for the two models are 0.85 and 1.02 seconds respectively. These are all increases to be expected as the networks are deeper, with more parameters and more convolutions, more recurrent processing, and more attention-based calculations. While these architectures tend to be more costly in terms of computing power, their forecasting accuracy and robustness is superior. Clearly, there is a dependency between model complexity and the number of computations required, as revealed by the overall trend. For more sophisticated models, longer training times are required, and the inference times are slightly higher. The testing times, however, are still not that high for all the architectures – even in the case of the most complex model, the testing time is still close to one second. This suggests that the hybrid solution proposed is well suited for practical application which makes real-time or near real-time forecasts of energy. The results bring an interesting trade-off in terms of accuracy and computational cost to the light of smart home IoT. Advanced models need extra training resources; however, training is usually carried out offline and only periodically. Once trained, the model would be able to make predictions very quickly, thus the extra time invested in training is worth it for the improved prediction capabilities. The small improvement in testing time proves that the proposed architecture provides an efficient operation while achieving better feature extraction, temporal learning, and attention-based optimization. The overall graph indicates that, while the hybrid model is more costly in terms of computation than simpler models, the cost is not too high and is offset by the significant benefits in terms of prediction accuracy, robustness and generalization. Thus the framework is a good balance between forecasting accuracy and computational cost, which is suitable for smart energy management systems in the real world and smart-home environments with IoT.

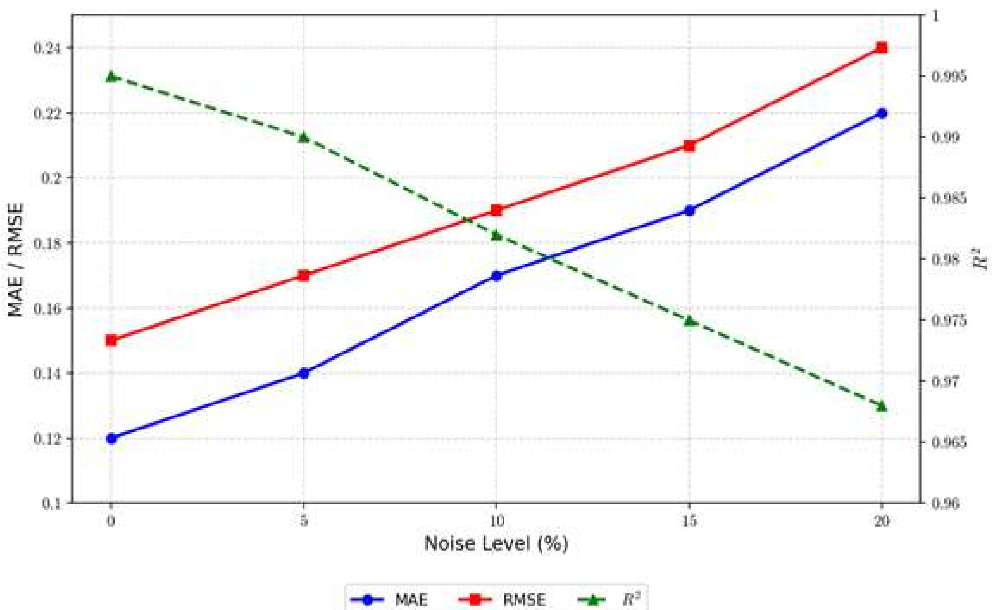

Figure 5 Noise Robustness Analysis

The graph in Figure 5 is a Noise Robustness Analysis which evaluates the performance of the proposed forecasting model for various level of artificial noise added to the input data. The noise levels range from 0% to 20% to simulate real world IoT environments with disturbances, errors in the transmitted data, missing data, or inaccuracies in the measured data. The model is tested for stability and resilience by plotting three important performance metrics – MAE, RMSE and R^2 with respect to the increasing noise. The blue MAE curve shows that the MAE curve increases in a gradual fashion with increasing noise intensity. The MAE value is found to be around 0.12 at 0% noise which shows that the prediction is very accurate. As the noise level increases to 5%, 10%, 15%, and 20%, MAE rises to approximately 0.14, 0.17, 0.19, and 0.22, respectively. This trend makes sense as bad input information decreases the accuracy of the model's consumption pattern identification. The increase is still relatively smooth and moderate, indicating that the proposed framework still has good prediction accuracy in the presence of noise. The curve of red RMSE is similar. RMSE increases from around 0.15 at 0% noise to approximately 0.17, 0.19, 0.21, and 0.24 as noise levels reach 5%, 10%, 15%, and 20%, respectively. The RMSE is a more severe (harsher) punishment for larger forecasting errors, thus the increase in RMSE with noise suggests that the model is effective in reducing the influence of large deviations. The without any sharp spikes of the RMSE further proves robustness of the proposed CNN–BiLSTM–Attention architecture. The green R2 curve, which is also shown on the secondary y-axis, shows that the R2 decreases with increasing noise. The model has a very high R2 value of ~0.995 at 0% noise, indicating that almost all of the energy consumption variation is captured by the model. As noise increases, R^2 decreases gradually to 0.991, 0.984, 0.975, and 0.968 at 5%, 10%, 15%, and

20% noise, respectively. Despite the decrease, the R^2 is still very high at the highest noise level, which means that the model still reflects the major temporal and consumption trends in the data. As seen in the overall behaviour of the graph, the proposed model exhibits good noise tolerance and generalization capabilities. The CNN block learns to extract well-distinguished local features from noisy input sequences, whereas the BiLSTM block is able to learn long-term temporal relationships even in the presence of data disturbances, and the Attention block emphasizes on informative patterns while suppressing the impact of irrelevant noisy signals. This means that the performance drop is not sudden but happens gradually. In practical applications, this robustness is very appealing for smart-home IoT energy forecasting systems where data generated by the sensors often include noise from the limitations of the sensor hardware, data delays, environmental noises, and data loss during transmission. The model shows relatively low error values and high R2 score even when noise is contaminated up to 20% assuring its applicability in real life energy management system. Overall, it validates that the suggested CNN – BiLSTM – Attention forecasting model is stable and reliable in noise augmentation. The model displays a gradual increase in MAE and RMSE and a gradual decrease in R^2 with the increase of noise, but the model still shows strong predictive capability for all test noise levels, which proves excellent model resilience and robustness for practical forecasting environments of energy consumption in IoT.

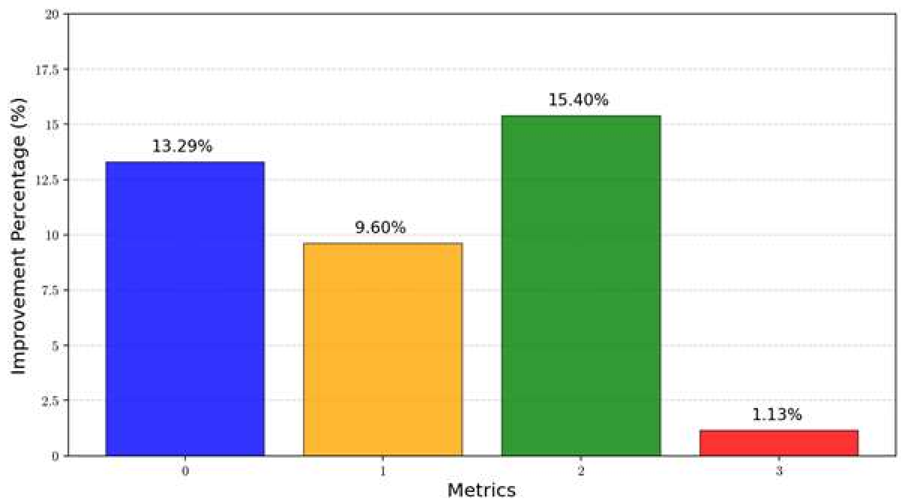

Figure 6 Performance improvement percentage of proposed Model vs CNN-LSTM

The figure 6 shows the Performance Improvement Percentage of the Proposed CNN–BiLSTM–Attention Model vs. CNN–LSTM baseline model for four evaluation metrics namely MAE, RMSE, MAPE, and R^2. The bars indicate the percentage increase when the model was implemented in research over the benchmark model. The first bar corresponds to a 13.29% decrease in the Mean Absolute Error (MAE). A lower MAE will mean more accurate forecasting as MAE is the average of the magnitude of the errors of the forecasts. The decrease from 0.143 to 0.124 represents that the proposed framework produces prediction values closer to actual values of the energy consumption resulting in better prediction precision. The second bar shows that the Root Mean Square Error (RMSE) was reduced by 9.60%. RMSE is a good measure of the robustness of the model, because it assigns larger penalties for prediction errors that are larger. The lower value of 0.179 from 0.198 indicates that the proposed architecture is capable of reducing the major forecasting errors, which results in more stable and reliable forecasting. The MAPE (Mean Absolute Percentage Error) is the highest improvement in the third bar, which is 15.40%. MAPE measures the prediction error as a percentage, and is widely used in energy forecasting applications as it is an intuitive measure of forecasting accuracy. The decrease of the relative prediction error from 4.87% to 4.12% proved that the proposed model could considerably reduce the relative prediction error and could be applied well in smart-home energy management systems in practice. The fourth bar shows an improvement in the coefficient of determination (R^2) of 1.13%. The percentage increase seems to be less than the error-based metrics, but the R^2 values are already very close to the maximum of 1. The higher R2 value (from 0.971 to 0.982) indicates that the proposed model is able to account for a higher percentage of the variance in the energy consumption data, thus having a better predictive ability and fitting the observed patterns well. In general, the chart shows that the CNN–BiLSTM–Attention model improves all metrics every time compared to the CNN–LSTM model. The most remarkable improvement is seen in MAPE which sees a gain of 15.40%, followed by MAE (13.29%) and RMSE (9.60%) and the R^2 gets an excellent value of 0.982. The enhancements confirm the efficacy of incorporating CNN feature extraction, BiLSTM temporal dependency learning, and Attention-based feature weighting, which enable the model to learn the spatial characteristics of consumption and also the long-term temporal relationships between the items. Based on this, the proposed approach offers more accurate, robust, and reliable energy consumption forecasting model for a smart-home IoT setting.

5.1 Results Analysis

In this section, the experimental validation and performance analysis of the proposed Optimized Attention-Based CNN–BiLSTM Framework for smart home IoT energy consumption prediction system were presented. The focus of the experimental investigation was on learning of training behavior, quantitative forecasting accuracy, comparative benchmarking, ablation analysis for individual components, statistical analysis of results and practical application for intelligent energy management. Training behavior analysis showed that the model learning process had a stable convergence characteristics and was optimized successfully. The proposed framework showed smooth decrease of the training and validation loss which showed that the parameters were adapted well and the tendency of overfitting was reduced. Dropout regularization, hyperparameter tuning, and adaptive optimization helped to boost the stability of learning and enhance the generalization ability. Their quantitative prediction analysis showed that they could predict well with all the evaluation metrics. The proposed framework presented a very low prediction error values for MAE, MSE, RMSE and MAPE values, and a high value for Coefficient of Determination (R2R2). The results show the good representation of the electricity consumption with a degree of nonlinearity and a good learning of the temporal dependencies embedded in smart home IoT datasets. The proposed framework was benchmarked with the ANN, RNN, LSTM, GRU and CNN–LSTM architectures, which showed the superiority of the proposed framework. The traditional machine learning methods and the standalone recurrent models showed relatively poor results, due to a limited ability in feature extraction, representation of context, or due to the lack of an adaptive feature weighting mechanism. The proposed architecture proposed is able to improve the forecasting accuracy by integrating the convolutional feature learning, the bidirectional sequential modelling and the attention-driven contextual prioritization into a single forecasting framework. The effect of individual architectural components was further validated by the ablation study. Experimental results showed that CNN feature extraction, bidirectional temporal learning using BiLSTM, and bidirectional temporal attention by adaptive weighting are effective for improving prediction each in its own way. The hybrid model design was well validated in the complete integrated framework which yielded the best forecasting results. Forecasting residuals were subjected to statistical analysis, showing that there was less bias in the forecasts, a stable distribution of forecast errors and a better robustness of the model. The derived results show good agreement with the actual energy trajectory proving the feasibility of the presented forecasting approach for real-world applications. With regard to the application, the proposed framework has great practical significance in the context of smart home IoT energy management, smart-grid demand forecasting, coordination of renewable energy sources and sustainable residential energy optimization. The ability to accurately predict can help aid in the intelligent scheduling, proactive demand management, cost reduction and environmentally conscious electricity use. Overall, the experimental results confirm the efficacy, solidness and feasibility of the proposed CNN–BiLSTM–Attention framework for intelligent energy forecasting in smart home environments with an IoT paradigm.

6 CONCLUSION

The overall summary of the complete research work is given in this subsection. It presents the research problem: prediction of energy consumption in smart homes with IoT devices, and briefly reviews the reasons for the problems of previous forecasting approaches in the prediction of non-linear and temporal dependent electric energy data. The proposed Optimized Attention-Based CNN–BiLSTM framework, along with CNN feature extraction, attention-based contextual weighting and BiLSTM bidirectional temporal learning will be summarized. It will talk about the experimental validation with the dataset of the smart home energy and comment on the results obtained in terms of MAE, RMSE, MAPE and R2R^2R2 indexes. Finally, the conclusion paragraph will highlight the improved forecasting accuracy, robustness and applicability of the proposed framework in the field of intelligent residential energy management. Possible extensions and improvements for future work are discussed in this subsection. The paragraph will delve into potential future research directions, including exploring the use of transformer-based forecasting models, graph neural networks, and hybrid multi-attention architectures for improving contextual learning. It could offer integration with edge computing, federated learning and TinyML to enable low latency real-time deployments in distributed IoT environments. Explainable Artificial Intelligence (XAI) methods to enhance the interpretability and trustworthiness of deep learning predictions can also be discussed. Additional extensions can include integration with renewable energy forecasting systems, multi-building prediction systems, coordination with smart-grid with adaptation, and transfer learning across domains for energy forecasting.

REFERENCES

1. Ou Ali IH, Agga A, Ouassaid M, Maaroufi M, Elrashidi A, Kotb H. Predicting short-term energy usage in a smart home using hybrid deep learning models. Front Energy Res. 2024;12. doi:10.3389/fenrg.2024.1323357.
2. Binbusayyis A. Energy consumption prediction using modified deep CNN-Bi LSTM with attention mechanism. Heliyon. 2025;11(1):e41507. doi:10.1016/j.heliyon.2024.e41507.

3. Anan M. Occupant-aware energy consumption prediction in smart buildings using LSTM networks. Energies. 2024;17(24):6451. doi:10.3390/en17246451.
4. Hammerschmitt BK. Deep learning for residential electrical energy consumption forecasting. Energies. 2025;18(22):5885. doi:10.3390/en18225885.
5. Severiche-Maury Z. LSTM networks for home energy efficiency. Designs. 2024;8(4):78. doi:10.3390/designs8040078.
6. Severiche-Maury Z. Forecasting residential energy consumption with the use of LSTM neural networks. Energies. 2025;18(5):1247. doi:10.3390/en18051247.
7. Severiche-Maury Z. Predicting energy consumption and time of use of home appliances using LSTM networks. Sustainability. 2025;17(11):4749. doi:10.3390/su17114749.
8. Badhe NB, Neve RP, Yele VP, Abhang S, Dhule KM, Mali D. An optimized system for predicting energy usage in smart grids using temporal fusion transformer and Aquila optimizer. Front Artif Intell. 2025;8:1542320. doi:10.3389/frai.2025.1542320.
9. Luo S. Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for short-term electricity load forecasting. Energy Rep. 2024;12:2676–2689.
10. Wen X. Deep learning-driven hybrid model for short-term load forecasting in smart grids. Sci Rep. 2024;14. doi:10.1038/s41598-024-63262-x.
11. Palan VA. Electricity consumption prediction using an advanced spatial-temporal GRU model. Sci Rep. 2026. doi:10.1038/s41598-026-46825-y.
12. Hasan MW. Design of an IoT model for forecasting energy consumption using improved LSTM. Internet Things Cyber-Phys Syst. 2025.
13. Im J. Data pipeline for real-time energy consumption prediction and management. Front Big Data. 2024;7. doi:10.3389/fdata.2024.1308236.
14. Liu C. Intelligent forecasting of electricity consumption for hospital energy management. Front Energy Res. 2025. doi:10.3389/fenrg.2025.1566664.
15. Yang X. An optimized method for short-term load forecasting based on time and weather-fused features using ConvLSTM-3D. Front Energy Res. 2025. doi:10.3389/fenrg.2024.1501963.
16. Byeon H. Deep learning and smart energy-based lightweight urban power load forecasting model. Front Sustain Cities. 2025. doi:10.3389/frsc.2024.1487109.
17. Shah SAH. Improved electric load forecasting using quantile long short-term memory network with dual attention. Energy Rep. 2025.
18. Balakrishnan A. Stacked hybrid model for load forecasting: integrating artificial neural networks, time series transformers and fuzzy logic transform. Sci Rep. 2025;15. doi:10.1038/s41598-025-04210-1.
19. Alsalem K. A hybrid time series forecasting approach integrating fuzzy clustering and machine learning for electricity demand prediction. Sci Rep. 2025;15. doi:10.1038/s41598-025-91123-8.
20. Zhou S. Adaptive electricity consumption forecasting approach for universal environments. Sci Rep. 2025;15. doi:10.1038/s41598-025-10147-2.
21. Tang H. Prediction of electricity consumption and hydropower generation using improved GRU and multi-feature self-attention. Sci Rep. 2025;15. doi:10.1038/s41598-025-32294-2.
22. Hussain A. Charging stations demand forecasting using LSTM-based encoder-decoder transformer model. Sci Rep. 2025;15. doi:10.1038/s41598-025-20421-y.
23. Shawon SM. Hybrid CNN-LSTM model for urban energy load forecasting with explainable artificial intelligence. Results Eng. 2025.
24. Li C. A novel CNN-LSTM-based forecasting model for electricity load prediction using CEEMDAN, self-attention, and autoencoder. Energy. 2025.
25. Li F. Optimal planning for integrated electricity and heat systems using CNN-BiLSTM-Attention forecasting. Energy. 2024.
26. Feng Z. A day-ahead building load forecasting method based on CNN-BiLSTM-Attention. Energy Build. 2026.
27. Luo S. Mid-term power load forecasting using an ensemble deep learning model with BKA optimization and CWGAN-GP data augmentation. Sci Rep. 2026. doi:10.1038/s41598-026-49674-x.
28. Yin Q. A review of research on building energy consumption prediction models. Sustainability. 2024;16(17):7805. doi:10.3390/su16177805.
29. Misiurek K. Review of methods and models for forecasting electricity consumption. Energies. 2025;18(15):4032. doi:10.3390/en18154032.
30. Taktak M. Evaluating the impact of frequency decomposition techniques on LSTM-based household energy consumption forecasting. Energies. 2025;18(10):2507. doi:10.3390/en18102507.
31. Zhang Y, Liu X, Wang J. Attention-based BiLSTM model for short-term household electricity demand forecasting. Energy Rep. 2024;10:2145–2158.