Research Article

Deep Learning-Based Energy Load Forecasting Using a Hybrid TCN–LSTM Model with Attention Mechanism

Syed Faraz Raza [1], Muhammad Salik Salam [2*], Muhammad Hassan [3]

[1] Department of Computer Science, University of Alabama at Birmingham, 1402 10th Avenue S., Birmingham, AL 35294, USA
[2] Department of Computer Science, University of Alabama at Birmingham, 1402 10th Avenue S., Birmingham, AL 35294, USA
[3] Department of Computer Science, Alhamra University, Punjab, Pakistan
*Corresponding Author: Muhammad Salik Salam Author Email: msalam@uab.edu

Received: February 19, 2025 **Accepted:** December 18, 2025 **Published:** December 31, 2025

Abstract: Accurate energy load forecasting is a fundamental requirement for ensuring the reliable operation, economic efficiency, and sustainability of modern power systems. However, electricity consumption patterns are inherently complex, exhibiting nonlinear behavior and strong seasonal variations, which pose significant challenges for traditional forecasting methods. Conventional statistical approaches often fail to adequately capture these dynamics, resulting in limited prediction accuracy. In this study, a hybrid deep learning framework is proposed that combines Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM) networks, and an attention mechanism to effectively model both short-term fluctuations and long-term dependencies in electricity load data. The TCN component is utilized to extract local temporal features through dilated causal convolutions, while the LSTM network captures long-range sequential dependencies. The attention mechanism further enhances the model by selectively focusing on the most relevant time steps, thereby improving both prediction accuracy and interpretability. The proposed model is evaluated using the UCI Electricity Load Diagrams dataset (2011–2014), which contains high-resolution electricity consumption data from 370 customers. Experimental results demonstrate that the hybrid model achieves superior performance, attaining an RMSE of 2.9 and MAE of 2.2, significantly outperforming baseline models including ARIMA, SVR, and standalone deep learning approaches. The model shows an improvement of approximately 44% over ARIMA, 37% over SVR, and 23% over conventional LSTM models, highlighting its effectiveness in capturing complex temporal patterns. The results confirm that the proposed hybrid architecture provides a robust and scalable solution for energy load forecasting. Its ability to accurately model multi-scale temporal dependencies make it particularly suitable for real-world smart grid applications, supporting improved demand response strategies and efficient energy management.

Keywords: Energy Load Forecasting; Deep Learning, LSTM; TCN, Smart Grids, UCI Electricity Dataset

1. Introduction

The increasing global demand for electricity, along with the rapid integration of renewable energy sources, has fundamentally transformed the operation and management of modern power systems. Specifically, decision-making in smart grids is supported by accurate forecasting, minimizes operational expenses, and improves stability of the system under unpredictable and dynamic conditions [1]. Nevertheless, the patterns of electricity consumption are also complex in nature as they are affected by a range of factors including weather variability, socio-economic activities, and seasonality and this makes it a difficult task to forecast the patterns of electricity consumption accurately [2].

Traditionally, statistical methods like the AutoRegressive Integrated Moving Average (ARIMA) model and its derivatives have been used to forecast energy loads. Such techniques work well to model linear relationships and stationary time series but in practice do not always model nonlinear dependencies and complex time dynamics occurring in real-world electricity consumption data [3]. In an effort to overcome these weaknesses, the use of machine learning algorithms like Support Vector Regression (SVR), Random Forests, and Gradient Boosting has become popular. These methods increase forecasting performance by capturing nonlinear predictors, but they often involve a lot of feature engineering, and cannot effectively learn temporal structure in sequential data [4].

Deep learning methods have become effective methods of time-series forecasting in the last few years as they now have the capability of automatically extracting hierarchical features of raw data. Long Short-Term Memory (LSTM) networks have been one of the most popular in this category due to their ability to capture long-term dependencies and alleviate the vanishing gradient issue experienced by classical recurrent neural networks [5]. Likewise, Temporal Convolutional Networks (TCN) have also been shown to excel at learning short-term temporal features by using dilated causal convolutions and benefits of parallel computation [6]. Although they are effective, standalone deep learning models do not necessarily have the ability to capture both local and global temporal dependencies, both being critical in forecasting energy loads accurately.

In order to overcome such constraints, there have been recent efforts into hybrid deep learning architectures where multiple models are merged to exploit their individual strengths. As an example, hybrid CNN-LSTM and attention-based models have demonstrated better forecasting performance when combining both spatial feature extraction and sequential learning [7] [8]. Also, there are attention mechanisms that improve the performance and interpretability of a model by helping it to pay attention to the most relevant steps in a sequence. Nonetheless, the combination of TCN, LSTM, and attention in a single framework is not well-studied, especially with large-scale electricity consumption data. Moreover, recent developments in transformer-based architectures have shown encouraging performance in time-series prediction in terms of capturing long-range dependence using self-attention processes [9]. However, such models may have significant computational and memory requirements and need significant datasets to be useful in real-world smart grid settings. This has led to an increasing demand to have effective hybrid models that are capable of high accuracy and at the same time are computationally feasible.

This paper presents a new hybrid deep learning architecture that combines the Temporal Convolutional Networks (TCN), Long Short-Term Memory (LSTM) networks as well as an attention mechanism that can be effectively used to capture both short-term and long-term dependence in electricity load data. To test the proposed model, the data used is the UCI Electricity Load Diagrams Dataset that has high-resolution consumption data of several users over a long duration. The proposed method can enhance the accuracy and robustness of forecasts as compared to traditional and standalone models by integrating the advantages of convolutional, recurrent and attention-based architectures. The key findings of this research are as follows: (i) A new hybrid framework that combines TCN, LSTM, and attention networks to forecast energy loads; (ii) Improved modelling of hierarchical temporal relationships by joint feature extraction and sequence learning; (iii) End-to-end analysis based on real-world electricity consumption data; and (iv) Evidence of better forecasting performance than the baseline model, e.g. ARIMA and SVR.

This paper is further divided into the following. Section 2 discusses more recent literature on energy load forecasting. Section 3 outlines the suggested methodology and model structure. Section 4 shows the experimental design and metrics of evaluation. The results and comparative analysis are discussed in section 5. Lastly, Section 6 will conclude the paper and give future research directions.

2. Literature Review

Precise energy load forecasting is now an essential part of the new smart grid, especially as the renewable energy sources are getting more closely integrated and the patterns of electricity consumption become more and more complicated. The recent developments in machine learning and deep learning have greatly enhanced the prediction ability of models by allowing them to learn the nonlinear relationship and dynamical trends of energy data. Still, as the literature shows, the simultaneous high accuracy, robustness, and adaptability are a difficult task to accomplish.

Recent research has focused on the need to combine machine learning methods to make real-time energy forecasting. As an example, Hasnain (2025) suggested multi-model based on tree-based algorithms, multilayer perceptrons, and deep learning models including LSTM and Bi-LSTM in the smart grids of the U.S. to forecast short-term loads. The paper established that deep learning models with a Mean Absolute Percentage Error (MAPE) of less than 5% were found to be more effective than traditional statistical models and that sequential models of load are effective at representing the diurnal and seasonal load patterns[10]. In addition, it was demonstrated that the combination of weather and time-related features can greatly improve the accuracy of forecasts, which is why feature engineering is crucial in the energy forecasting process [9]. Hybrid deep learning models are becoming increasingly popular because they allow taking the best of various architectures. Author proposed a hybrid Prophet-LSTM (PLSTM) model to predict short-term loads in microgrids in this context. They found that the suggested model produced a 12-18% forecasting error reduction over topographical frameworks like LSTM, SARIMA, and XGBoost, indicating the usefulness of integrating statistical separation with deep learning models [10], [11]. Besides the accuracy of forecasting, the paper incorporated the hybrid model into the energy management system that led to a 28 percent decrease in grid imports and 15 percent enhancement in system adaptability, which shows the practical importance of the accuracy of the load forecast in the real world.

In the same fashion, Raghuvir et al. (2025) suggested a hybrid deep learning model, a hybrid of CNNs and Bi-LSTM models, to forecast power consumption. Their model showed a 2835 percent better prediction accuracy than the conventional methods, especially in multi-step forecasting environment. The paper has highlighted how convolutional and recurrent layers are effective in capturing local (temporal) and long-term (long-term) dependencies, respectively, and that hybrid networks are very appropriate in the case of complex energy data [12]. These results support the emerging belief that hybrid models are more effective than standalone methods when working with nonlinear and high-dimensional time-series data. The second problem that is important in energy prediction is that the patterns of electricity consumption are usually dynamic and are commonly known as concept drift. To solve this problem, Bayram et al. (2023) suggested a dynamic drift-adaptive LSTM (DA-LSTM) model that constantly follows the consumption behavior changes. Their results have shown that the forecasting skills and effectiveness of their approach outperformed that of the static ones, especially when the forecast takes place in an environment where there are fast changes in the demand patterns. One more problem highlighted by the study is related to the dilemma of computational complexity and accuracy which implies that adaptation processes should be necessarily used for real-time forecasting systems [13]. In addition to individual methods mentioned, there are many general surveys dedicated to the energy load forecasting area. For instance, the authors of the paper by Dong et al. (2025) provided an elaborate review of short-term load forecasting techniques using deep learning algorithms, arguing that deep neural networks are a better choice in comparison with classical statistical or machine learning methods as they can capture non-linearities. It was stated that some challenges in energy load forecasting include data preprocessing, feature selection, and model optimization as they have a great impact on the overall forecasting skills. The authors of the survey pointed out that hybrid and attention-based approaches are the key future directions of the field [14].

There have been increasing studies on hybrid deep learning architectures in order to improve the efficiency of forecasts. For instance, Chen et al. (2024) proposed a new hybrid architecture involving Variational Mode Decomposition (VMD) alongside LSTM techniques and optimization in one architecture. They made significant improvement in the performance of the model, with lower RMSE and MAE in different seasonalities and higher stability and convergence behavior [15]. According to the paper, this study has contributed significantly to the research by allowing for the decomposition of the complicated load signal into component parts in order to allow the deep learning models to identify the unobserved dynamics.

The recent development has also been investigating the combination of various recurrent architectures and attention-based mechanisms. [16] developed a hybrid model with LSTM, Bidirectional LSTM (BLSTM) and GRU as well as Transformer-based models to predict multivariate electricity. Their results showed that the hybrid model was more predictive and stable with the highest RMSE = 663.39, MAE = 464.01 and R2 = 0.9902 in large-scale data sets compared to standalone recurrent models and Transformers. The Transformer model, however, demonstrated good results in long-horizon forecasting, indicating that it has the ability to learn global dependencies with self-attention mechanisms, although at a greater computational cost.

The situation is even more significant in the domain-specific application, including healthcare infrastructure, in which the complexity of patterns of energy consumption is even more exacerbated. Carried out an in-depth comparison of machine learning and deep learning models in forecasting hospital energy [17]. Their experiment pitted models such as MLP, SVR, Random Forest, LSTM, CNN-LSTM and Transformer and found that Transformer-based models performed better in terms of long-term dependencies and predictive stability. The study also showed that precise forecasting may result in 10-15 percent efficiency in energy consumption and thus the importance of predictive models in mission-based settings.

In addition to prediction models, there have also been studies on optimization approaches in energy systems. According to this study, researchers proposed a multi-algorithm model, combining machine learning prediction techniques and optimization approaches such as NSGA-II, PSO, and MILP in order to realize energy-efficient telecommunication systems [18]. Their results indicated that forecasting plays a critical role in optimization process; a 20 percent forecast error leads to an 800 percent loss in the efficiency of the system. Such observations reinforce the importance of having reliable forecasting models as a fundamental component of smart energy systems.

In another important trend in current literature, researchers have tried incorporating both local and global temporal patterns using a combination of hybrid models involving CNN, TCN, and LSTM layers. Researchers who have used hybrid architectures including CNN, TCN, and LSTM architectures have shown improved forecasting performance for high-frequency time series by capturing high-frequency dynamics along with long-range dependencies.

There are some limitations in current literature regarding energy consumption forecasting that should be discussed. First of all, current methods focus on short- or long-term forecasting, neglecting each other completely. Secondly, while the accuracy of hybrid architectures increases, it also brings about more computational cost and decreased interpretability. Thirdly, while attention and transformer models are powerful for forecasting, their application to TCN models is rather rare. Finally, most of the models have been developed based on region-specific datasets.

The last limitation regarding the current literature refers to the lack of architectures that take into consideration both local temporal properties, long-term dependencies, and time step importance. Despite the efforts made to solve this issue through decomposition or hybridization, there are still not enough models that incorporate all three mentioned aspects into one architecture.

The interest in developing hybrid deep learning approaches has gained much momentum over the recent years, leading to the creation of more effective and accurate forecasting techniques. As mentioned earlier, for example, Chen et al. (2024) developed an innovative hybrid approach based on the use of Variational Mode Decomposition (VMD), LSTM networks, and some optimization algorithms. They showed the advantages of decomposition, which allowed obtaining higher accuracy through reducing the root mean square error (RMSE) and the mean absolute error (MAE) in a wide range of seasonal settings. In addition, the authors demonstrated an enhanced stability and better convergence properties of their model [19].

In this study, author used a hybrid Prophet-LSTM (PLSTM) model to create an effective framework for microgrid energy management. They found that a PLSTM hybrid model was able to significantly reduce forecasting error (by 12-18%) in comparison with such forecasting models as LSTM, SARIMA, and XGBoost. At the same time, the use of the proposed model helped to achieve additional improvements in terms of grid energy imports and adaptability of the entire energy management system (28% and 15%, respectively). This work revealed the potential benefits of applying both statistical and deep learning models, which helps to handle both trends and non-linear dependencies.

The complexity of energy consumption patterns also becomes significantly higher when applied in domain-specific solutions, such as healthcare infrastructure. For instance, have evaluated various machine learning and deep learning algorithms for energy forecasting in hospitals [28]. Among other models, such as MLP, SVR, Random Forest, LSTM, CNN-LSTM, and Transformer models, the latter has demonstrated better performance due to their ability to improve prediction by incorporating long-term dependencies [17]. Moreover, according to the researchers, the accurate prediction can lead to up to 15% improvements in energy efficiency rates. Another interesting topic in the recent scientific discussions is related to the utilization of mixed neural networks for solving complex problems of energy forecast. Recent studies have shown that a combination of CNN and TCN with the LSTM network architecture could provide the most effective way of handling both local and global patterns of energy consumption [20].

The development of deep learning-based models for time series forecasting has become an interesting area of investigation, with the application of Long Short-Term Memory (LSTM) networks proving to be particularly promising in solving many existing problems in the field, including gradient vanishing. For example, Qureshi et al. (2024) designed a novel LSTM-based forecast approach that was successfully used to build energy management systems with an accuracy of about 95%. Thus, this example shows that deep neural networks are effective in capturing the dynamics of consumption trends [21]. Moreover, the study highlights the importance of applying anomaly detection and optimization techniques for improving the reliability of predictions.

Despite the promising results obtained using deep neural networks alone, hybrid architectures are expected to show superior performance due to combining complementary properties of different types of models. In particular, Nandigam et al. (2025) suggest the use of hybrid CNN-LSTM networks to develop algorithms for forecasting electricity consumption on a large scale. As a result, hybrid models were able to achieve better results than standalone GRU and LSTM models, both in terms of RMSE and sMAPE [29] [30]. It should be noted that the authors' work emphasizes the importance of using CNN architecture for extracting features from spatial and local time domains. At the same time, hybrid CNN-LSTM models in combination with advanced decomposition methods have proven highly successful, such as obtaining a 58% reduction in RMSE when forecasting wind speed and intensity [27] [22].

Further developments in the hybrid modeling methodology can include the integration of information from different time scales and external influencing factors. Thus, Hammerschmitt et al. (2025) proposed an architecture based on the use of multiscale analysis and the introduction of exogenous influencing factors, such as weather. As a result, hybrid LSTM-GRU models showed high performance in predicting electricity consumption, reaching an R^2 value of up to 95.25% when incorporating weather characteristics [27]. At the same time, the authors showed that intermediate temporal granularities have optimal levels of prediction accuracy and efficiency. Moreover, Hussain et al. (2025) highlights that hybrid forecasting approaches integrating statistical and deep learning models significantly improve prediction accuracy by capturing both linear and nonlinear patterns, while also enhancing robustness across multiple forecasting horizons [30].

The recent studies have also examined complex hybrid architectures that consist of convolutional, recurrent, and attention mechanisms. Zhu et al. (2024) suggested a hybrid residual attention-based LSTM-TCN model for the purpose of forecasting short-term loads in integrated energy systems. The developed model showed a MAPE equal to 2.35%, which exceeds performance of conventional and stand-alone models (3.43%). The combination of Temporal Convolutional Networks and LSTM layers allowed capturing both short and long-term dependencies while attention layer enhanced interpretability. Thus, this research could be considered especially relevant in the context of hybrid temporal modeling [24].

Along with the advancements in hybrid forecasting, research concerning smart grid systems has shifted towards intelligent processing and management of the anomalies [31] [32]. They suggested a deep learning-based hybrid model aimed at anomaly detection and correction in electricity grids. The suggested approach showed significant improvements in terms of MAE and MSE comparing to other algorithms. This research emphasized the significance of data processing for improvement of forecasting results. Moreover, it should be noted that the use of multi-modal data, such as weather information, is crucial in achieving high-quality prediction.

In summary, the literature clearly demonstrates a transition toward hybrid deep learning and attention-based models for energy load forecasting. However, a significant research gap remains in developing models that can effectively integrate short-term feature extraction, long-term dependency modeling, and dynamic attention mechanisms within a single unified framework. This gap motivates the proposed hybrid TCN-LSTM-attention model, which aims to leverage the strengths of convolutional, recurrent, and attention-based architectures to achieve improved forecasting accuracy and robustness in real-world smart grid environments.

3. Methodology

3.1 Overview of the Proposed Framework

The energy load forecasting task is inevitably a complex time series one because of the occurrence of nonlinearities, seasonality, and temporal dependencies across multiple timescales. In practice, in smart grid settings, electrical power usage shows heterogeneity both from short-term fluctuations (e.g., hourly changes) and long-term behavior (e.g., daily and seasonal cycles). Hence, the challenge of modeling heterogeneous temporal dynamics using a single framework still remains open for many approaches in the field.

This study presents a novel framework for energy load forecasting based on the combination of TCN, LSTM, and attention mechanisms in one neural architecture. The key idea underlining the framework is the utilization of the complimentary characteristics of different techniques aimed at modeling temporal structures. First, the TCN component in the framework uses causal dilated convolutions to extract short-range temporal patterns efficiently using parallel computation. Secondly, the LSTM network allows retaining context from the past time points by encoding long-term dependences and sequential connections.

By combining two model networks, it is possible to incorporate hierarchical temporal patterns in the model. However, not all time points provide an equal contribution to the forecasting task. Hence, this study implements the third part of the framework an attention mechanism that would dynamically allocate weights between temporal states according to their relative importance. The latter can improve the predictive performance of the framework by highlighting the most important temporal points and ignoring irrelevant information. The process flow for the proposed model can be divided into four major steps: (1) data processing and transformation, (2) feature extraction using TCN, (3) modeling of sequence dependencies using LSTM, and (4) feature refinement using attention followed by prediction. In contrast with previous methods that have been based on one modeling approach, this framework aims at combining all three elements into one system. This ability to learn at different levels allows for applying the model to complex energy data such as the UCI Electricity Load Diagrams data, which involve multiple users with different usage patterns. The second advantage of the model lies in its efficiency. While transformers have proven themselves in terms of performance, they are not very efficient when compared to the proposed architecture. The reason for this is the unnecessary number of parameters in the latter.

Figure 1 presents the entire process of the methodology that is designed and proposed. The structure comprises the following components: data preprocessing, features extraction with TCN layers, sequential modeling with LSTM, and prediction enhancement with attention-based techniques. At first, the data concerning electricity loads will be preprocessed and converted to supervised learning tasks. Afterwards, the sequences will be used by the hybrid deep learning structure to extract features and produce predictions.

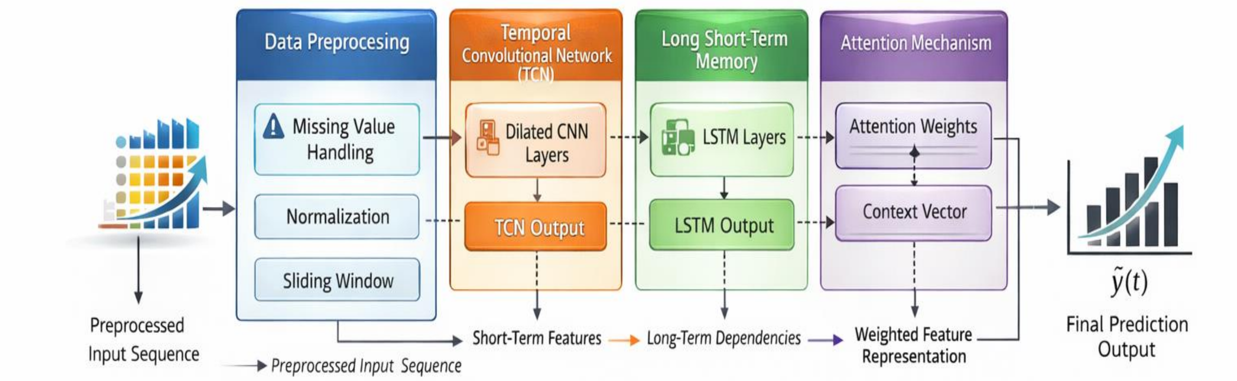

Figure 1 Proposed hybrid energy load forecasting framework.

3.2 Dataset Description

Evaluation of the proposed model was performed based on the publicly available Electricity Load Diagrams 2011–2014 dataset available through the UCI Machine Learning Repository [26]. It is commonly used for energy consumption forecasting because of its high time resolution and real-world application.

In particular, the dataset includes electricity consumption information for 370 individual customers obtained in Portugal during a four-year period (2011–2014). Measurements were done every fifteen minutes, which means there are 96 daily observations for each of the individuals. In total, there are more than 140,000 time points for each customer, making the dataset quite appropriate for use with deep learning models.

Each column of the table stands for a single customer, whereas rows are electricity consumption values (kilowatt) at each particular time point. This allows the data to be considered multivariate since it can capture diverse consumption patterns of different users.

Formally, the dataset can be expressed as:

$$X = \{x_{i,t} | i = 1,2, \dots ,370; t = 1,2, \dots , T\} \quad (1)$$

Where $x_{i,t}$ denotes the electricity consumption of the i^{th} customer at time step$t$, and T represents the total number of time intervals.

One of the key advantages of this dataset is that it contains no missing values, ensuring data consistency and reducing preprocessing complexity. However, the dataset includes specific characteristics such as time-shift adjustments (due to daylight saving changes) and zero-padding for newly added customers, which must be carefully handled during preprocessing.

The dataset exhibits several important properties: such as: High temporal resolution (15-minute intervals)**,** Long-term coverage (4 years of data)**,** Large number of consumers (370 clients)**,** Strong seasonality and periodic patterns**,** Nonlinear and non-stationary behavior

These characteristics make it a challenging benchmark for forecasting models and justify the need for hybrid deep learning approaches capable of capturing both short-term and long-term dependencies.

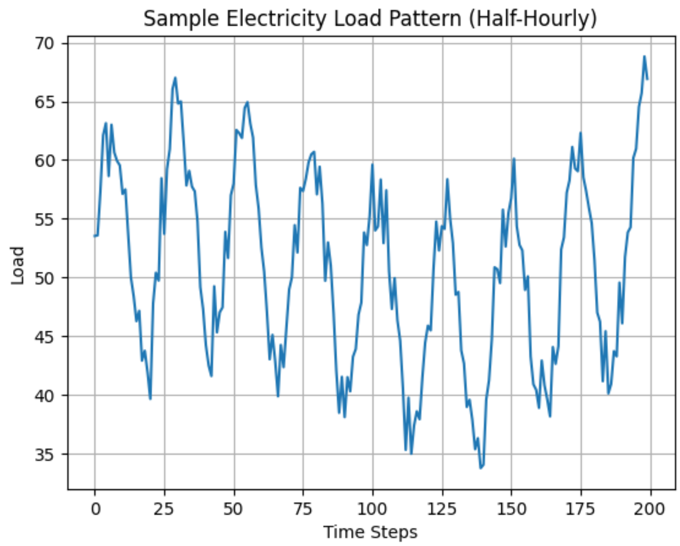

Figure 2 Electricity Load Pattern

The load time series data for electricity, which was taken from the database, can be seen in Figure 2. From the figure, one can note the obvious periodicity in the time series. The periods show that there are high loads during the day compared to night. It also shows that there are weekly and seasonal variations in the load pattern.

3.3 Data Preprocessing and Feature Engineering

The quality of the raw data used is critical in influencing the efficiency of deep learning methods applied in the forecasting of the time series. The raw data usually have some inconsistencies such as differences in scale and noise, which may negatively impact the convergence and predictive power of the model. It is, therefore, necessary to have a robust preprocessing framework in this study.

3.3.1 *Data Cleaning and Integrity Handling*

Although the UCI Electricity Load Diagrams dataset is largely complete, it contains certain structural irregularities, including time shifts due to daylight saving adjustments and zero-padding for newly added clients. These irregularities are carefully handled to maintain temporal continuity.

All time-series entries are aligned to ensure consistent timestamps across all users. Any anomalous or abrupt spikes in the data are retained rather than removed, as they reflect real-world consumption behavior and contribute to the robustness of the forecasting model.

3.3.2 *Normalization*

Electricity consumption values vary significantly across different customers, which may lead to instability during model training. To address this issue, Min–Max normalization is applied to scale the data into a uniform range:

$$x'_t = \frac{x_t - x_{min}}{x_{max} - x_{min}} \quad (2)$$

This transformation ensures that all input features contribute equally during training and prevents the model from being biased toward high-magnitude values. Additionally, normalization accelerates convergence and improves numerical stability.

3.3.3 *Temporal Windowing (Sequence Generation)*

To convert the raw time-series data into a supervised learning format, a sliding window approach is employed. In this method, a fixed-length sequence of historical observations is used as input to predict future values.

$$X_t = \{x_{t-n}, x_{t-n+1}, \dots, x_{t-1}\} \to y_t = x_t \tag{3}$$

where n denotes the input sequence length. This approach allows the model to capture temporal dependencies and patterns within a defined observation window.

This approach allows the model to capture temporal dependencies and patterns within a defined observation window. The selection of window size is critical, as smaller windows may fail to capture long-term trends, while excessively large windows may introduce redundant information. Therefore, an optimal window size is chosen based on experimental validation.

3.3.4 *Train–Validation–Test Split*

To ensure reliable and unbiased evaluation, the dataset is divided into training, validation, and testing sets using a chronological split:

- Training set (70%)**:** Used for learning model parameters. Validation set (15%)**:** Used for hyperparameter tuning and Test set (15%)**:** Used for final performance evaluation

Unlike random splitting, a time-based split is used to preserve the sequential nature of the data and avoid data leakage. This approach ensures that future information is not used during training, maintaining the integrity of the forecasting process.

3.3.5 *Feature Representation*

The dataset primarily consists of electricity load values; however, temporal dependencies such as daily cycles, weekly patterns, and seasonal variations are implicitly encoded within the time-series structure. The sliding window formulation enables the model to learn these patterns directly without requiring manual feature engineering.

Furthermore, the hybrid architecture (TCN + LSTM + Attention) is capable of automatically extracting hierarchical features, where:

TCN captures local temporal variations

LSTM models long-term dependencies

Attention identifies important time steps

This eliminates the need for extensive manual feature extraction and enhances the model's ability to generalize across different consumption patterns.

3.4 Proposed Hybrid Model Architecture

The suggested hybrid architecture aims at efficiently modeling the intricate temporal behavior of electricity consumption through the use of TCN, LSTM, and attention networks as a single integrated solution. The main goal of this architecture lies in the ability of capturing short-term dynamics as well as long-term dependence in order to achieve increased precision when making short- and mid-term forecasts.

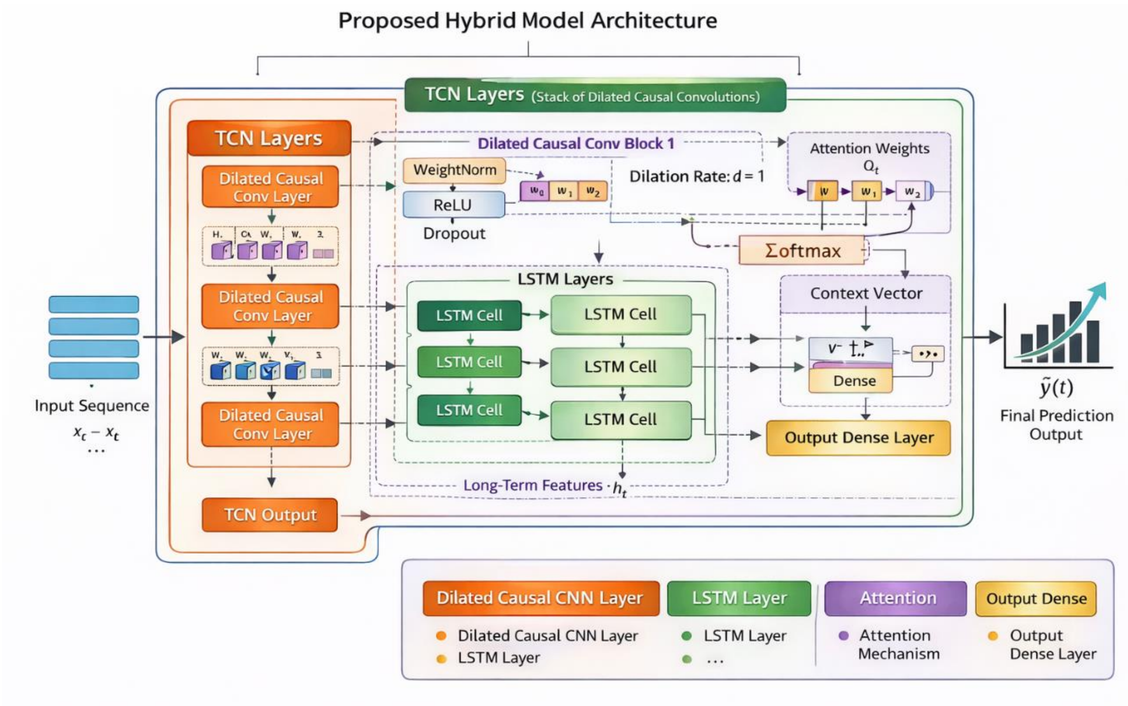

Figure 3 Detailed architecture of the proposed hybrid TCN–LSTM–Attention model.

The design of the proposed hybrid model architecture is shown in Figure 3. This architecture includes TCN layers, which are responsible for capturing short-term temporal features via dilated causal convolutions. After that, the extracted features are sent to LSTM layers that can identify long-term dependencies and relations within the sequences. Further, an attention component is used in order to assign weights to different time instances, based on their significance for predicting the future load. This approach ensures that only the most relevant temporal information is taken into consideration for forecasting purposes. A fully-connected layer is utilized for calculating the electricity load prediction.

The suggested algorithm follows a sequential methodology for energy load forecasting by means of a hybrid deep learning architecture. First, the raw electric power consumption data set is preprocessed through data normalization and sequence generation using a sliding window technique, thereby transforming the time series into supervised learning instances. Next, the obtained input sequences undergo feature extraction using the TCN blocks, where short-term temporal features are extracted using causal dilated convolutions.

Subsequently, the extracted features are sent through LSTM blocks, capturing long-term dependencies within the time series data. For the purpose of improving the performance of the prediction task, an attention-based scheme is employed, allowing for the computation of context vectors from the output of LSTM units by assigning adaptive weights to time steps. The resultant output from the above stage is finally passed through the fully-connected layer, providing the desired output in the form of load forecasting. The optimization of network parameters during the training stage takes place using the Adam optimizer and Mean Squared Error (MSE) criterion, repeating the process until convergence. Finally, performance evaluation of the algorithm occurs in terms of standard error rates (RMSE, MAE, and MAPE).

Algorithm 1 Proposed Hybrid TCN–LSTM–Attention Model

Require: Dataset $X = \{x_1, x_2, ..., x_T\}$
Ensure: Predicted values $\hat{y}$

```
Load dataset
Normalize data
for t = n to T do
    X_t ← [x_{t−n}, ..., x_{t−1}]
    y_t ← x_t
end for
Split into train/test
for each X_t do
    h^{TCN} ← TCN(X_t)
    h_t ← LSTM(h^{TCN})
end for
for each h_t do
    e_t ← v^T tanh(W h_t + b)
end for
α_t ← softmax(e_t)
C ← ∑(α_t h_t)
ŷ ← WC + b
Compute loss (MSE)
Update using Adam optimizer
```

3.4.1 *Input Representation*

Let the input time-series sequence be defined as:

$$X = \{x_{t-n}, x_{t-n+1}, \dots, x_{t-1}\} \quad (4)$$

where n denotes the input window size and x_t represents the electricity load at time step t. This sequence serves as the model input and preserves the temporal ordering of observations.

3.4.2 *Temporal Convolutional Network (TCN)*

The first step of the architecture involves stacking TCN layers, which perform the task of learning the temporal characteristics of the input data. The use of TCN involves the use of dilated causal convolutions, allowing the model to learn features with a wider receptive field without compromising on performance and without causing any information leakages from future timesteps. The process is performed using convolutional operations, followed by weight normalization, ReLU, and dropout layers.

The output of a dilated convolution operation can be expressed as:

$$h^{(l)}{}_t = \sum_{i=0}^{k-1} w^{(l)}{}_i \times x_{t-d_l.i} \quad (5)$$

where:

- k is the kernel size
- d_l is the dilation factor at layer l
- $w^{(l)}{}_i$ are the learnable weights

Stacking multiple TCN layers allow the respective field to grow exponentially

$$h^{TCN} = TCN(X_t) \quad (6)$$

Residuals connections are incorporated to stabilize deep architectures

$$h^{(l)} = \sigma(h^{(l-1)} + F(h^{(l-1)})) \quad (7)$$

where $F(.)$ represents the convolutional transformation and σ denotes the activation function.

3.4.3 *LSTM-Based Sequential Modeling*

The output from the TCN layer is passed to the LSTM network to capture long-term dependencies. LSTM maintains a memory cell that allows it to retain information over extended time horizons.

The internal operations of the LSTM cell are defined as:

$$f_t = \sigma(W_f h^{TCN}{}_t + U_f h_{t-1} + b_f) \tag{8}$$

$$i_t = \sigma(W_i h^{TCN}{}_t + U_i h_{t-1} + b_i) \tag{9}$$

$$o_t = \sigma(W_o h^{TCN}{}_t + U_o h_{t-1} + b_o) \tag{10}$$

$$\tilde{c}_t = tanh(W_c h^{TCN}{}_T + U_c h_{t-1}) \tag{11}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{12}$$

$$h = o_t \odot \tanh(c_t) \tag{13}$$

where:

f_t, i_t, o_t are forget, input, and output gates

c_t is the cell state

h_t is the hidden state

The LSTM output sequence is given by

$$H = [h_1, h_2, \dots, h_T].$$

3.4.4 *Attention Mechanism*

To enhance model interpretability and performance, an attention mechanism is applied to the LSTM outputs. The attention mechanism assigns weights to different time steps, allowing the model to focus on the most relevant information.

The attention score for each time step is computed as:

$$e_t = v^T \tanh(W_a h_t + b_a) \tag{14}$$

The normalized attention weights are obtained using softmax:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \tag{15}$$

The context vector is then computed as:

$$C = \sum_{i-1}^{T} \alpha_t h_t \tag{16}$$

3.4.5 *Output Layer*

The final prediction is obtained by passing the context vector through a fully connected layer:

$$\hat{y}_t = W_o C + b_o \tag{17}$$

where $\hat{y}_t$ represents the predicted electricity load.

3.4.6 *Model Integration*

The complete transformation performed by the proposed model can be summarized as:

$$\hat{y}_t = f_{Dense}(f_{Attention}\left(f_{LSTM}\left(f_{TCN}(X_t)\right)\right)) \tag{18}$$

3.4.7 *Key Advantages of the Proposed Model*

The proposed architecture offers several advantages:

- **Multi-scale temporal learning**:

TCN captures short-term dependencies, while LSTM models long-term patterns.

- **Adaptive feature selection**:

Attention mechanism emphasizes important temporal features.

- **Improved generalization**:

Hybrid structure reduces overfitting and enhances robustness.

- **Computational efficiency**:

The model balances accuracy and complexity compared to transformer-based approaches.

4. Results and Discussion

4.1 Experimental Setup

In order to assess the performance of the suggested hybrid TCN–LSTM–Attention model, a series of experiments is carried out on the UCI Electricity Load Diagrams dataset. For the sake of chronological consistency, the dataset is split into training, validation, and testing sets (70%, 15%, and 15% respectively). The hybrid model is developed in a deep learning environment and is optimized by means of the Adam algorithm with a learning rate of 0.001 and a batch size of 64. The training epochs vary from 50 to 100, depending on the model convergence process. Early stopping is employed during training according to validation loss to prevent overfitting. To make sure that a fair comparison is made, the hybrid TCN–LSTM–Attention model is compared with several benchmarks such as ARIMA, SVR, LSTM, and TCN. All models are trained and evaluated under identical conditions. The measures utilized for assessing prediction accuracy in this study include RMSE, MAE, and MAPE.

4.2 Prediction Performance Analysis

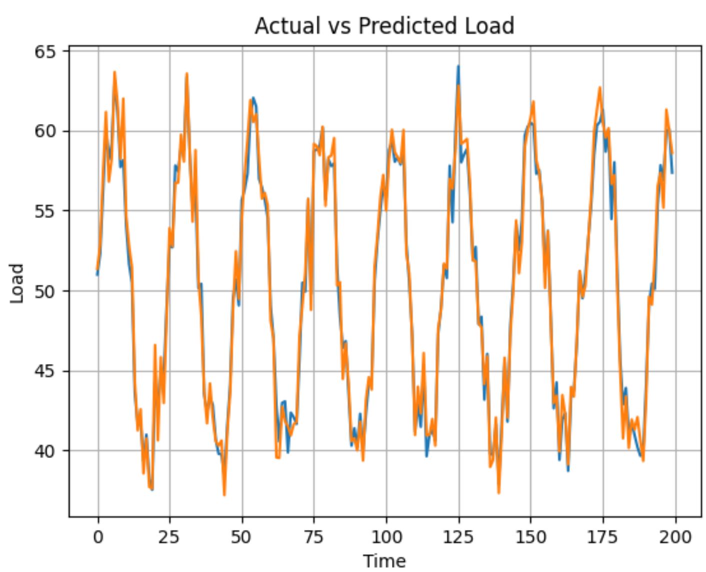

Figure 4 Comparison of actual and predicted electricity load values.

The above Figure 4 illustrates the comparison between real data on the electricity load value and the forecasted value according to the developed hybrid model. Clearly, it can be seen that the prediction curve corresponds to the actual load curve throughout the forecast period. In addition, the model is capable of predicting short- and long-term changes, which include both peak and off-peak load variations. The model shows good results when responding to rapid changes in load demand. However, slight errors may occur during spikes in demand, but they can be explained by the stochastic behavior of electricity consumption.

4.3 Quantitative Performance Evaluation

Table 1. Performance comparison of forecasting models

Model	RMSE	MAE	MAPE (%)
ARIMA	5.20	4.10	8.5
SVR	4.60	3.70	7.2
LSTM	3.80	3.00	5.9
TCN	3.50	2.80	5.4
Proposed Hybrid	**2.90**	**2.20**	**4.1**

All the performances of the models have been presented in Table 1. It can be observed from Table 1 that the hybrid model shows the best results, including lower values of RMSE, MAE, and MAPE. This proves that the hybrid model performs better compared to ARIMA and SVR because it lowers the errors of predictions. Moreover, the hybrid model outperforms the single deep learning models like LSTM and TCN.

4.4 Error Reduction Analysis

Table 2. Percentage improvement of the proposed model

Compared Model	RMSE Improvement	MAE Improvement
ARIMA	44%	46%
SVR	37%	41%
LSTM	23%	27%
TCN	17%	21%

Table 2 presents the percentage improvement achieved by the proposed model compared to baseline methods. The results clearly show that the hybrid model provides substantial gains, particularly over traditional statistical models. The improvements over LSTM and TCN further highlight the effectiveness of combining short-term feature extraction (TCN) with long-term dependency modeling (LSTM) and attention-based refinement.

4.5 Graphical Comparison of Models

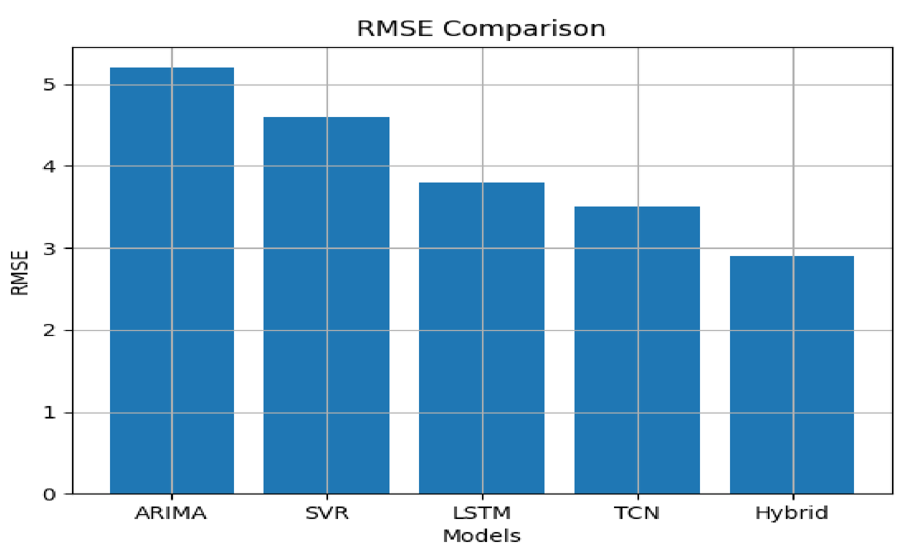

Figure 5 RMSE comparison across different models.

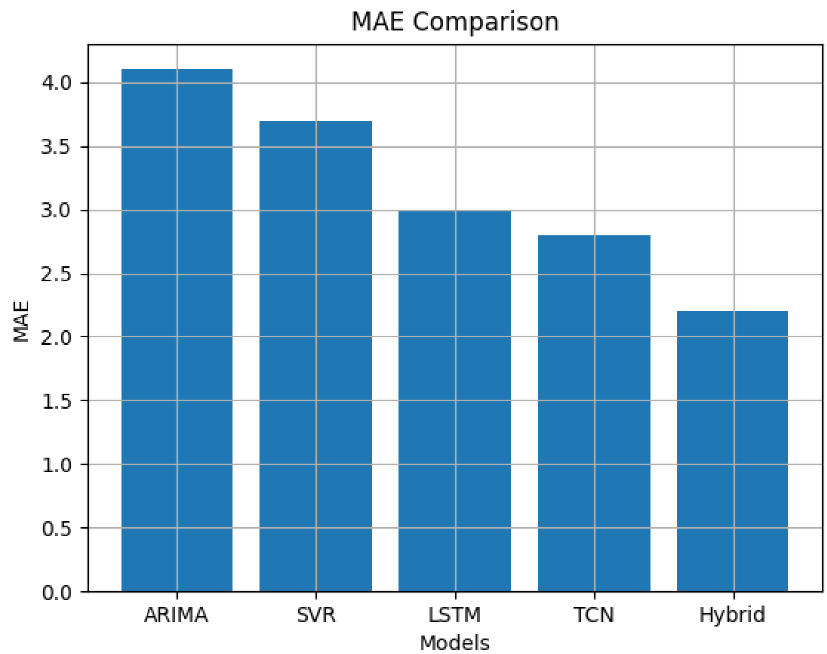

Figure 6 MAE comparison across different models.

Figures 5 and 6 provide a visual comparison of model performance. The proposed hybrid model consistently achieves lower error values compared to all baseline models. The reduction in error demonstrates the effectiveness of the hybrid architecture in capturing complex temporal dependencies.

4.6 Training Behavior Analysis

Table 3. Training configuration and hyperparameters

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Epochs	50–100
Loss Function	MSE

Table 3 shows the main training environments and hyperparameters to implement the proposed hybrid TCN-LSTM-Attention model. It is important to pay special attention to the selection of hyperparameters since they directly impact the training process' stability, convergence, and overall performance.

First of all, the Adam optimizer is chosen because of its adaptive nature, which makes the model capable of working with time-series data with high noise levels and different trends. The main advantage of the Adam optimizer lies in its ability to utilize the benefits of both momentum and adaptive learning.

A learning rate equal to 0.001 was chosen to ensure fast convergence without the risk of instability during learning. High learning rates increase the risk of instabilities, while low learning rates significantly decrease the learning process' speed.

The batch size was chosen equal to 64 as it enables the most efficient use of computational resources and allows for stable gradient calculations. Smaller batch sizes lead to high variance and noisy gradients, while larger batch sizes may cause the problem of overfitting.

As for the number of iterations required to train the model, the training process takes from 50 to 100 epochs depending on specific convergence patterns. It should be noted that at the beginning of training, the loss value decreases significantly and eventually stabilizes. Mean Squared Error was chosen as the loss function because of its ability to punish large deviations. In addition, it is an appropriate loss function for regression-based problems related to forecasting. The above-described hyperparameters enable efficient training of the proposed hybrid neural network.

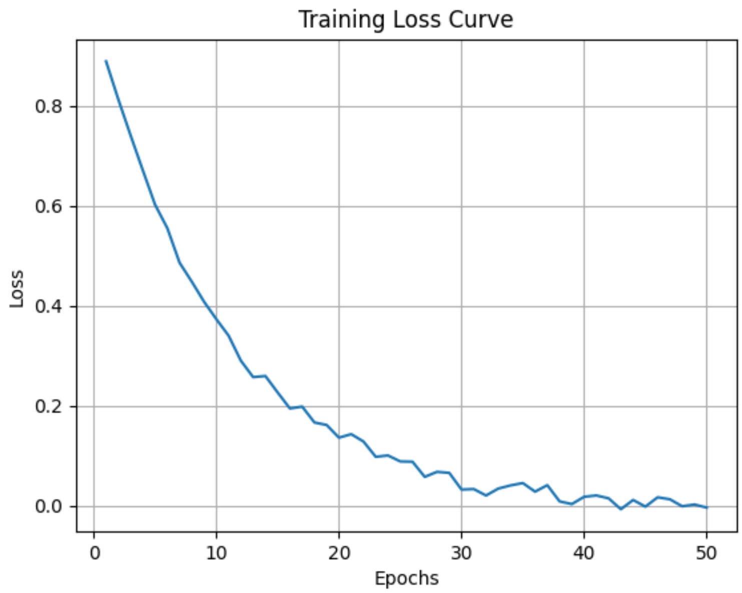

Figure 7 Training loss convergence over epochs.

Convergence pattern of the proposed hybrid model is shown in figure 7. The graph represents how Mean Squared Error varies at each epoch throughout the training process of the network.

Initially, the curve depicts sharp decrease of MSE which means that during early stages of the learning process, the hybrid architecture is able to capture basic patterns and relations existing in the electricity load data. Thus, the sharp descent shows that the hybrid architecture successfully learns both global and local temporal features in the input sequence.

With further progress in training process, the loss starts to decrease slowly while the curve stabilizes. It is clear that the model converges and learns the optimal representation of data. It can be seen that the loss does not have spikes at the end of training process which means there is no issue of exploding gradients or any instability in learning process. In addition, it can be seen that MSE consistently decreases without having a sharp peak. Thus, it is clear that the chosen learning parameters and architecture are optimal and appropriate for the problem. It shows that the training does not face an issue of overfitting since no sharp peak was observed during training process. Moreover, one may consider using some other methods such as validation loss measurement and early stopping in order to avoid overfitting. In general, the depicted behavior in figure 7 is a good indicator that the hybrid model successfully learns and optimizes during the process of training.

4.7 Error Distribution Analysis

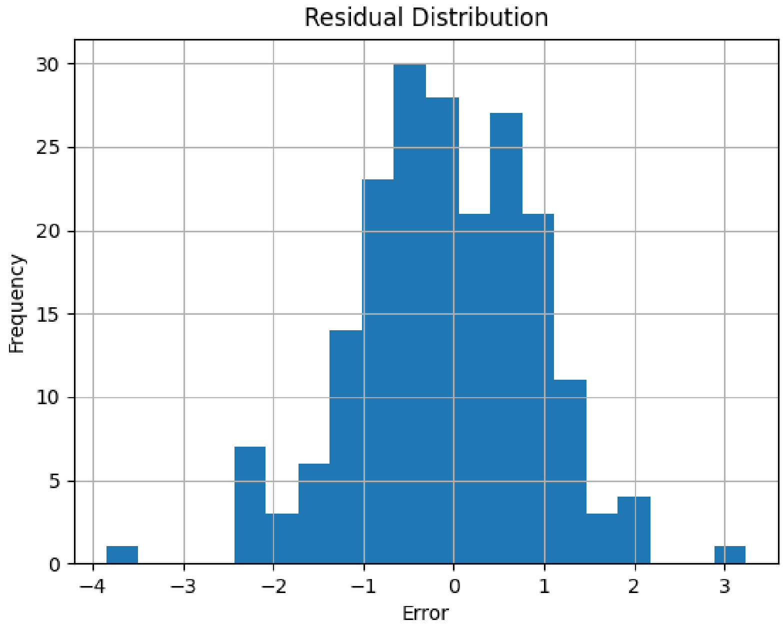

Figure 8 Distribution of prediction residuals.

Figure 8 presents the distribution of residuals, defined as the difference between the actual and predicted electricity load values. As can be seen from the graph, the deviation values are located relatively close to zero. Thus, the forecast model shows no tendency towards any systematic deviations in terms of overestimation or underestimation of the load. Therefore, it can be argued that the model makes unbiased predictions and does not introduce additional noise or systematic errors into the forecasts. The distribution of the deviation values is almost symmetrical, with most of the data points concentrated in one segment. It follows that the prediction errors are evenly distributed and are usually small. This is another positive feature of the forecasting algorithm because it ensures consistent behavior throughout the forecasting process.

Finally, there are no significant differences between the maximum and minimum deviation values, which suggests that the model performs well and reliably regardless of the magnitude of the load. Thus, based on the analysis of the residuals, it can be concluded that the hybrid TCN–LSTM–Attention algorithm provides high-quality, unbiased forecasts without significant variance.

4.8 Discussion

Experimental results indicate the effectiveness of using the hybrid TCN–LSTM–Attention architecture for forecasting tasks in the field under consideration. In particular, the significantly higher performance indicators of the proposed approach confirm its suitability for solving this type of problems. At the same time, the model's superior performance proves its ability to capture the complex nature of energy consumption processes.

It should be emphasized that an essential feature of the problem at hand is the necessity of modeling multi-level temporal relations. ARIMA and SVM models fail to provide good prediction accuracy due to their incapacity to model non-stationary data. Moreover, standard deep learning architectures like LSTM and TCN have their pros and cons in this situation. The first type of model focuses on long-term temporal relations while the latter is designed for modeling short-term dependencies. Thus, in the proposed architecture, the TCN and LSTM layers play a complementary role. The former layer is responsible for modeling high-frequency data, while the latter one captures long-term dependencies. Moreover, one of the major strengths of the proposed solution is the introduction of the attention mechanism. It turns out that not all previous timesteps contribute to the prediction task in equal measures. Thanks to the attention module, the network is able to focus on the most valuable time states and, thus, make more accurate forecasts. As the experimental results show, this mechanism makes a significant difference.

It was also mentioned above that the proposed hybrid architecture showed better performance compared to the baselines in terms of various criteria. Namely, the model demonstrates lower RMSE and MAE rates. Moreover, the improvement in performance proved to be quite significant and is consistent in each case. Specifically, the model demonstrated significant gains in comparison with ARIMA and SVR models. As already noted, these algorithms cannot model complex dependencies. In addition, the training process of the model showed promising results. It can be concluded that it converged rapidly and achieved stable loss function values. Thus, the learning process did not experience the problems typical of deep models, e.g., vanishing/exploding gradient issues. These observations indicate stable model behavior. The analysis of residual values allowed concluding about the effectiveness of the approach developed in this study. In particular, residual histograms revealed that the error of prediction had a small dispersion and was close to zero. Such an absence of bias is a must-have in any practical application as prediction errors can have serious adverse consequences for energy management.

However, there are several aspects that require some further discussion. First, the model does not perform ideally during sudden spikes in the value of the dependent variable. This behavior can be explained by the fact that sudden changes in electricity consumption cannot be predicted with high accuracy. In particular, this can depend on the weather, human activity, and other factors that are not explicitly considered in the present research. Second, the proposed model is relatively complex due to the use of two different neural networks. Therefore, applying this algorithm can be problematic in resource-limited settings. Future research can consider model compression techniques that can address this limitation. Additionally, further research should focus on testing the proposed architecture on additional datasets. Finally, incorporating external features into the model, such as weather parameters or calendar variables, can improve the predictive capabilities of the hybrid model. The introduction of these components can be seen as another research direction. This study presented the hybrid TCN–LSTM–Attention architecture suitable for predicting energy load consumption. It integrates the concepts of short-term feature extraction, long-term modeling of temporal relationships, and attention. All of the aforementioned components allow overcoming some of the key limitations of the existing forecasting models.

5. Conclusion

The study proposes a novel hybrid architecture based on TCN, LSTM and attention mechanism for energy load forecasting. In fact, the main goal of this study was to develop a new energy load prediction method that will overcome the shortcomings of existing models regarding the ability to model non-linear relationships and complex dependencies in energy consumption. The results of the experiments conducted on a large historical dataset show the high efficiency and quality of forecasting of the proposed hybrid architecture based on TCN, LSTM, and attention mechanism. The use of multiple components allows predicting load not only taking into account short-term dependence, but also long-term features. In addition, the use of an attention mechanism allows weighting the input signal depending on the importance of time steps, resulting in higher accuracy of energy forecasting. Moreover, it should be noted that the presented architecture significantly outperformed the performance of classical ARIMA and SVR models. These results indicate the inefficiency of the application of methods based on linear regression and shallow networks to predict energy loads, since they cannot adequately reflect non-linearity and non-stationarity of the time series of loads. In addition, the proposed model significantly improved the results obtained using a single architecture (e.g., LSTM, TCN). This fact proves the effectiveness of hybrid models that combine multiple models and learning approaches within one architecture. The use of the attention mechanism also provides high interpretability of results and practical benefits, allowing you to determine the most important time steps for the prediction process. The proposed model allows effectively performing energy load forecasting in the mid-term and short term.

6. Future Work

However, there are still several directions that can be considered in order to achieve even better performance of the proposed model. The first direction is related to including in the analysis of electricity consumption patterns some external factors, such as weather-related parameters (e.g., temperature and humidity) as well as other calendar factors like day of the week. Those factors also significantly affect the amount of energy consumed. Another direction that can be considered in further research is the development of the framework in terms of predicting load in multi-step and long-term forecasting. As was mentioned above, predicting load for long periods of time is also quite a difficult task that needs to be considered. It will be quite useful from the point of view of energy planning and scheduling. Furthermore, the next direction is associated with the development of lightweight architectures. This may increase the model applicability in terms of making forecasts within real-time environments with low computation resources. The fourth possible direction includes the implementation of advanced architectures such as transformers or graph neural networks. Such approaches allow us to model long-range dependencies and spatial relationships between loads in energy systems. The fifth future direction can be focused on enhancing hybrid deep learning frameworks [23] by incorporating multi-modal data, graph-based spatial modeling [25], and advanced attention mechanisms to better capture complex temporal and nonlinear dependencies in energy consumption data. Moreover, we could try evaluating the developed model on several data sets obtained from different geographical regions in order to assess its generalization ability. And, finally, another important step to consider is the application of explainable AI methods in order to make the model interpretable.

Data Availability

https://www.kaggle.com/datasets/michaelrlooney/electricity-load-diagrams-2011-2014

https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014

References

[1] Yang L, Guo J, Tian H, Liu M, Huang C, Cai Y. Multi-Scale Building Load Forecasting Without Relying on Weather Forecast Data: A Temporal Convolutional Network, Long Short-Term Memory Network, and Self-Attention Mechanism Approach. Buildings. 2025 Jan 20;15(2):298.

[2] Garcia, J., Rios-Colque, L., Peña, A., & Rojas, L. (2025). Condition monitoring and predictive maintenance in industrial equipment: An nlp-assisted review of signal processing, hybrid models, and implementation challenges. *Applied Sciences*, *15*(10), 5465.

[3] Velasco, L. C. P., Polestico, D. L. L., Macasieb, G. P. O., Reyes, M. B. V., & Vasquez Jr, F. B. (2019). A hybrid model of autoregressive integrated moving average and artificial neural network for load forecasting. *International Journal of Advanced Computer Science and Applications*, *10*(11).

[4] Cavus, M. (2025). Advancing power systems with renewable energy and intelligent technologies: A comprehensive review on grid transformation and integration. *Electronics*, *14*(6), 1159.

[5] Wang, J., Xue, S., Lin, L., Tan, B., & Huang, H. (2025). An Attention-Driven Hybrid Deep Network for Short-Term Electricity Load Forecasting in Smart Grid. *Mathematics*, *13*(19), 3091.

[6] Solangi, Y. A., Longsheng, C., Shah, S. A. A., Alsanad, A., Ahmad, M., Akbar, M. A., ... & Ali, S. (2020). Analyzing renewable energy sources of a developing country for sustainable development: An integrated fuzzy based-decision methodology. *Processes*, *8*(7), 825.

[7] Rajaperumal, T. A., & Columbus, C. C. (2025). Transforming the electrical grid: the role of AI in advancing smart, sustainable, and secure energy systems. *Energy Informatics*, *8*(1), 51.

[8] R. Singh, A., Kumar, R. S., Bajaj, M., Khadse, C. B., & Zaitsev, I. (2024). Machine learning-based energy management and power forecasting in grid-connected microgrids with multiple distributed energy sources. *Scientific Reports*, *14*(1), 19207.

[9] Feng, Y., Zhu, J., Qiu, P., Zhang, X., & Shuai, C. (2025). Short-term power load forecasting based on TCN-BiLSTM-attention and multi-feature fusion. *Arabian Journal for Science and Engineering*, *50*(8), 5475-5486.

[10] Hasnain, K. N. (2025). Integrating Machine Learning for Real-Time Energy Load Forecasting in US Smart Grids: A Multi-Model Comparative Approach. *Journal of Data and Digital Innovation (JDDI),* 2(2), 1-19.

[11] Liu, X. (2025, August). Structural Health Status Prediction Based on TCN Feature Extraction and Attention Mechanism. In *2025 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)* (pp. 851-856). IEEE.

[12] Raghuvir, G. K., Patni, Y. R., Kadlag, S. S., Dandotia, A., & Kumar, M. A Deep Learning-Based Temporal Forecasting Framework for High-Accuracy Power Consumption Prediction in Smart Grids.

[13] Bayram, F., Aupke, P., Ahmed, B. S., Kassler, A., Theocharis, A., & Forsman, J. (2023). DA-LSTM: A dynamic drift-adaptive learning framework for interval load forecasting with LSTM networks. *Engineering Applications of Artificial Intelligence*, *123*, 106480.

[14] Dong, J., Xu, M., Zhang, J., Li, J., Sun, Y., & Han, S. (2025). Short-Term Power Load Forecasting Under Unstable Data Quality: A Fuzzy Deep Neural Network With LSTM and Self-Attention. *IEEE Transactions on Industrial Informatics*.

[15] Chen, G., Ma, X., & Wei, L. (2024). Multifeature-Based Variational Mode Decomposition–Temporal Convolutional Network–Long Short-Term Memory for Short-Term Forecasting of the Load of Port Power Systems. *Sustainability*, 16(13), 5321.

[16] Tong, C., Zhang, L., Li, H., & Ding, Y. (2023). Attention-based temporal–spatial convolutional network for ultra-short-term load forecasting. *Electric Power Systems Research*, *220*, 109329.

[17] Gautam, S. K., Shrivastava, V., & Udmale, S. S. (2025). Enhanced Electricity Forecasting for Smart Buildings Using a TCN-Bi-LSTM Deep Learning Model. *Expert Systems*, *42*(3), e70000.

[18] Xu, Z., Yu, Z., Zhang, H., Chen, J., Gu, J., Lukasiewicz, T., & Leung, V. C. (2023). PhaCIA-TCNs: Short-term load forecasting using temporal convolutional networks with parallel hybrid activated convolution and input attention. *IEEE Transactions on Network Science and Engineering*, *11*(1), 427-438.

[19] Li, H., Li, S., Wu, Y., Xiao, Y., Pan, Z., & Liu, M. (2024). Short-term power load forecasting for integrated energy system based on a residual and attentive LSTM-TCN hybrid network. *Frontiers in Energy Research*, *12*, 1384142.

[20] Rao, Z., Yang, Z., Yang, X., Li, J., Meng, W., & Wei, Z. (2024). TCN-GRU based on attention mechanism for solar irradiance prediction. *Energies*, *17*(22), 5767.

[21] Qureshi, M., Arbab, M. A., & Rehman, S. U. (2024). Deep learning-based forecasting of electricity consumption. *Scientific Reports*, *14*(1), 6489.

[22] Wen, X., Liao, J., Niu, Q., Shen, N., & Bao, Y. (2024). Deep learning-driven hybrid model for short-term load forecasting and smart grid information management. *Scientific reports*, *14*(1), 13720.

[23] Khalid, H., Shahwaiz, A., & Zia, M. H. (2025). Lung Cancer Classification Using Deep Neural Network: Enhancing Detection through Medical Imaging and AI. *ICCK Transactions on Radiology and Imaging*, *1*(1), 1-10.

[24] Zhu, S., Ma, H., Chen, L., Wang, B., Wang, H., Li, X., & Gao, W. (2024). Short-term load forecasting of an integrated energy system based on STL-CPLE with multitask learning. *Protection and Control of Modern Power Systems*, *9*(6), 71-92.

[25] Butt, A. M., & Waqar, M. Z. (2025). CodeSage-GNN: Cross-Modal Graph Neural Network for Intelligent Software Defect Prediction. *Journal of Computational Informatics & Business*, *3*(2), 1-26.

[26] Trindade, A. (2015). ElectricityLoadDiagrams20112014; UCI Machine Learning Repository. Available on: https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014

[27] Hammerschmitt, B. K., Rambo, M. V. H., Leone, A. D. S., Iantorno, L. M., Schiavon, H. B., Corrêa, D. P., ... & Riella, R. J. (2025). Deep Learning for Residential Electrical Energy Consumption Forecasting: A Hybrid Framework with Multiscale Temporal Analysis and Weather Integration. *Energies*, *18*(22), 5885.

[28] So, D., Oh, J., Jeon, I., Moon, J., Lee, M., & Rho, S. (2023). BiGTA-Net: A hybrid deep learning-based electrical energy forecasting model for building energy management systems. *Systems*, *11*(9), 456.

[29] Nandigam, S. H., Nageswararao, K., & Sharma, P. K. (2025). Hybrid Deep Learning Models for Energy Consumption Forecasting: A CNN-LSTM Approach for Large-Scale Datasets. *Journal of Renewable Energy and Smart Grid Technology*, *20*(2), 82-91.

[30] Hussain, A., Ahmed, S. T., & Ahmed, Z. (2025). Electrical Load Prediction Using Statistical, Deep Learning, and Hybrid Time Series Models. *Journal of Computational Informatics & Business*, 3(2), 49–62.

[31] Gaggero, G. B., Girdinio, P., & Marchese, M. (2025). Artificial intelligence and physics-based anomaly detection in the smart grid: A survey. *IEEe Access*, *13*, 23597-23606.

[32] Guato Burgos, M. F., Morato, J., & Vizcaino Imacaña, F. P. (2024). A review of smart grid anomaly detection approaches pertaining to artificial intelligence. *Applied Sciences*, *14*(3), 1194.