Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

Akıllı Şebeke Ortamlarında Dikkat Tabanlı Enerji Talep Tahmini

Yunus Emre ISIKDEMIR[1*], Fuat AKAL[2]

[1]Computer Engineering Department, Graduate School of Science & Engineering, Hacettepe University, Ankara, Türkiye.
[2]Computer Engineering Department, Hacettepe University, Ankara, Türkiye.
[1]yeisikdemir@gmail.com, [2]akal@hacettepe.edu.tr

Geliş Tarihi: 21.01.2024
Kabul Tarihi: 02.04.2024

Düzeltme Tarihi: 10.03.2024

doi: 10.62520/fujece.1423120
Araştırma Makalesi

Alıntı: Y.E. Işıkdemir ve F. Akal, "Akıllı şebeke ortamlarında dikkat tabanlı enerji talep tahmini", Fırat Üni. Deny. ve Hes. Müh. Derg., vol. 3, no 3, pp. 227-240, Ekim 2024.

Öz

Akıllı şebeke, modern enerji peyzajının kritik bir unsuru olup, artan enerji taleplerini karşılamak için güvenilir, verimli ve sürdürülebilir bir yol sağlamaktadır. Bununla birlikte, akıllı şebeke teknolojisi tarafından üretilen büyük miktardaki veri, gelişmiş veri işleme ve analiz tekniklerinin geliştirilmesini gerektirmektedir. Bu makalede, akıllı şebeke uygulamalarında zaman serisi tahmininde kullanılmak üzere, dilatasyonlu konvolüsyon ve dikkat mekanizmalarını birleştiren bir dikkat tabanlı zaman serisi iş akışı öneriyoruz. Bu akış, dilatasyonlu konvolüsyonları kullanarak zaman serisi verilerinden zamansal özellikler çıkarır ve dikkat mekanizmalarını kullanarak gizli durumlardaki önemli zaman noktalarını vurgular. Deneysel değerlendirmeler sonucunda, enerji talebi tahmininde, yaygın olarak kullanılan derin öğrenme tabanlı yöntemlere göre %8'e kadar daha iyi bir performans gösterdiği gözlemlendi. Bu kazancı diğer modellerin aldığı eğitim süresinin yalnızca 1/3'ü kadar bir sürede elde edilmiştir. Ayrıca, tamamen farklı bir alanda %42'lik bir kazanç elde edilmiştir ve akışın diğer alanlara uyarlanabileceği gösterilmiştir. Bu çalışma, araştırmacılara akıllı şebeke uygulamaları için daha doğru ve verimli tahmin modelleri geliştirmelerine yardımcı olabilir, ayrıca enerji sistemlerinin sürdürülebilir yönetimi ve akıllı şebeke operasyonlarının optimizasyonu için yapay zeka ve dikkat tabanlı tahmin tekniklerinin potansiyeli hakkında değerli bilgiler sunabilir.

Anahtar kelimeler: Akıllı şebeke, Zaman serisi tahmini, Attention mekanizması, Dönüştürücüler, Enerji talebi tahmini

*Yazışılan yazar

Firat University Journal of Experimental and Computational Engineering
FUJECE

Attention Based Energy Demand Forecasting in Smart Grid Environments

Yunus Emre ISIKDEMIR [1*], Fuat AKAL [2]

[1]Computer Engineering Department, Graduate School of Science & Engineering, Hacettepe University, Ankara, Türkiye.
[2]Computer Engineering Department, Hacettepe University, Ankara, Türkiye.
[1]yeisikdemir@gmail.com, [2]akal@hacettepe.edu.tr

Received: 21.01.2024
Accepted: 02.04.2024

Revision: 10.03.2024

doi: 10.62520/fujece.1423120
Research Article

Citation: Y.E. Işıkdemir and F. Akal, "Attention based energy demand forecasting in smart grid environments", Firat Univ. Jour.of Exper. and Comp. Eng., vol. 3, no 3, pp. 227-240, Octobor 2024.

Abstract
The smart grid is a crucial aspect of the modern energy landscape, providing a reliable, efficient, and sustainable way of meeting the growing energy demands. However, the vast amounts of data generated by smart grid technology necessitate the development of advanced data processing and analysis techniques. In this paper, we propose an attention-based time series workflow that combines dilated convolution and attention mechanisms for time series forecasting in smart grid applications. This workflow extracts temporal features from time series data using dilated convolutions and emphasizes significant temporal points in the hidden states using attention mechanisms. Experimental evaluations showed up to an 8% better performance for energy demand forecasting compared to commonly used deep learning-based methods. Our workflow achieved this gain by requiring 1/3 of the training time other models took. We also improved performance by 42% in various domains, demonstrating the adaptability of our approach across different areas. This study may assist researchers in constructing accurate forecasting models for smart grid environments. Furthermore, it highlights that the attention-based approach can be employed to promote sustainable energy and optimize smart grid environments.

Keywords: Smart grid, Time series forecasting, Attention, Transformers, Energy demand forecasting

*Corresponding author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

1. Introduction

A smart grid is a modern electrical power grid that includes advanced technologies such as digital communication, automation, and monitoring [1]. It is an effective and environmentally friendly power distribution system created to address increasing energy needs while keeping the environmental impact to a minimum level [2]. A smart grid can adaptively control the flow of electricity based on immediate shifts in demand and supply. The use of smart grids helps humanity for energy management, reduced costs, increased energy efficiency and more sustainable future. It helps incorporate renewable energy into the grid and allows energy to flow both ways between consumers and the grid [3,4].

Time series based energy demand forecasting plays an important role in smart grids [5]. Since precise energy demand forecasting may allow to optimize energy distribution, reduce energy waste and efficient energy management. In this way, companies supply energy based on expected demand and help to avoid energy shortages using the results of this forecasting. [6-8].

Deep learning is a type of machine learning that utilizes artificial neural networks to identify complex patterns in large data sets. It mimics the activation of the human brain and neurons using multiple nodes. Each of these nodes has weights to learn from the data and extract useful information [9]. With the use of Internet of Things (IoT) devices and sensors in the smart grid, large amounts of data are generated and necessitating the use of deep learning algorithms. The use of these algorithms enables the analysis of large data sets, such as energy consumption. This allows for the estimation of possible future consumption and optimization of the energy supply.

An attention mechanism, that has become quite popular recently and achieved successful results, is a concept of modeling the relationship between two sequences. It allows to focus specific part of the input sequence and produce specific output. In this study, this mechanism is used to capture dependencies between different time steps for accurately forecasting energy demand Moreover, this mechanism may also help to make the model more interpretable by identifying the key factors that influence the output [10].

The Encoder Decoder Long Short Term Memory architecture, also known as Seq2Seq, is a commonly employed framework for sequence modeling [11]. The model uses two neural networks which are the encoder and decoder. The encoder learns to extract features and represent the data in a latent vector form. The decoder learns to reconstruct the output sequence. However, the bottleneck issue limits efficient transmission of information in this architecture because there is a single connection between the encoder and the decoder [12]. This study used an attention mechanism to mitigate the bottleneck issue in this framework. This was achieved by establishing multiple connections from the encoder to the decoder. The connections prioritize the most important time steps. By doing so, the attention mechanism may facilitate the efficient transmission of information and improve the overall performance of the model.

The subsequent sections of this paper are structured as follows: related works which focus on relevant literature; methodology which provides a detailed research design, data collection and analysis procedures; experimental results which include the findings of the statistical analyses; and conclusion which summarizes the main findings of the study.

2. Related Works

With the use of smart grid technology becoming more and more common, there is a lot of data that needs advanced analysis techniques. Making accurate predictions about the future based on time series data is really important for smart grid operations. Since it helps us understand how people use energy and allows us to manage energy systems more efficiently. Therefore, many researchers conduct experiments to create good models for estimating the future trends in energy demand. In this section, we will provide an overview of the current state of forecasting methods in smart grids.

Chujai and colleagues [13] used the Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models to predict how much power individual households would use each day, week, and quarter. The study found that the ARIMA model worked best for forecasting power usage on a monthly and quarterly basis. On the other hand, the ARMA model was found to be more effective for shorter time periods, like predicting daily and weekly power consumption. These results are significant for managing energy consumption because they offer guidance on the best methods for predicting power usage based on different timeframes.

In their study, Cascone et al. [14] suggested a two step method for predicting household energy usage. They used a mix of Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM) architecture. In the first step, they applied an LSTM model to estimate the total active power generated over a 500-hour period. For the second step, they employed a combination of convolutional neural network features and LSTM for a weekly energy consumption forecast. The outcomes of the study were positive, showing potential benefits for predicting power consumption.

Seliverstova et al. [15] highlight that electricity consumption forecasting can be divided into four types based on the timeframe, namely ultra-short-term, short-term, medium-term, and long-term. The study employed various algorithms, including ARIMA, Group Method of Data Handling (GMDH), LSTM, and seq2seq. After conducting experiments, it was found that GMDH performed the best among the tested methods.
Tae-Young Kim and Sung-Bae Cho [16] employed CNN-LSTM hybrid deep learning approach for forecasting power consumption in their study. CNN layer was utilized to simplify the complexity of the spatial information and extract the most important features for forecasting. LSTM layer was applied to model the sequential relationship. Deep Neural Network was utilized to interpret the results and generate the forecasted value. They achieved highly accurate power consumption forecasts.

Shi and Wang [17] proposed a combination of landmark based spectral clustering (LSC) and deep learning techniques to cluster and forecast the power consumption dataset. The dataset was transformed into a matrix, and missing values were imputed. The data samples were then divided into three distinct clusters based on their periodicity and regularity using LSC.

Keskin et. al. [18] introduced an approach that utilizes a hierarchical architecture of LSTM (HLSTM). The proposed method is tested on real life crime, electric power consumption, and financial data sets. Results show that the HLSTM improves modeling of distant temporal interactions compared to traditional LSTM architecture.

Oh and Lee [19] used dense sampling method to capture more information from the input data. The proposed method samples data from both the time and window axes, resulting in a larger training dataset and other benefits such as model-agnosticism and easier window selection.

3. Methodology

In this study, we propose a time series workflow that is capable of capturing both long range and short range relationship of the data, shown in Figure 1. The workflow consists of data understanding and analysis, dilated convolution operation for temporal feature extraction, LSTM with attention for learning sequential correlation, and forecasting steps. The analysis step involves downsampling the dataset [20], into daily intervals and imputing missing values using spline interpolation. Temporal features are then extracted using dilated convolution and fed to LSTM layers. To prevent bottleneck, hidden states of all LSTM layers are used instead of a single connection of LSTM layers. The attention layer is then used to focus on the most important parts of the series to map output using an encoder-decoder LSTM architecture. The information is concatenated and the final prediction is performed. This workflow offers a comprehensive approach to time series forecasting that effectively incorporates multiple techniques and processes to improve the accuracy and reliability of predictions.

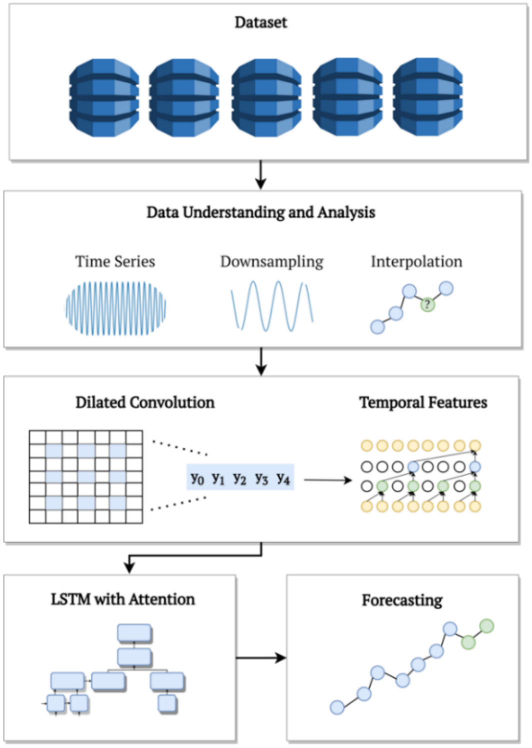

Figure 1. Workflow of the proposed methodology

The primary objective of this study is to determine the effectiveness of the proposed workflow for a forecasting model, which can accurately predict future energy consumption patterns. This accuracy will enable proper planning and management of electricity distribution, thereby contributing to the efficient and sustainable use of energy resources. To further validate our approach we conducted our experiments on two other time series datasets and one of them was from a different domain. The subsequent subsections will provide comprehensive details on the aforementioned steps.

3.1. Dataset

The Hourly Energy Demand Generation and Weather Dataset is a collection of hourly measurements of electricity generation, demand, and weather conditions. The dataset consists of observations from 2015 to 2018. Each row represents the measurement for each hour of the day. As our primary dataset, we used it to analyze the relationship between energy demand and weather patterns, which is crucial for predicting future energy consumption and ensuring the stability of the smart grid. The dataset includes a wide range of weather variables such as temperature, wind speed, and precipitation, as well as energy demand and generation data from various sources such as solar, wind, and fossil fuel power plants. In the context of this study, the variable of interest is the total load, which represents the actual electrical demand. The aim of this research, however, is to develop benchmark models for comparing the effectiveness of our proposed approach to other existing models in predicting the total load variable, which is crucial in various fields including energy management, weather forecasting, and infrastructure planning. Therefore, we considered two other datasets to further validate our approach.

First, the Individual Household Electric Power Consumption Dataset provides information on the electricity consumption of a single household for four years, from December 2006 to November 2010. It includes data points for power consumption within the household collected every minute. The dataset contains 2,075,259 records. The features include the date and time of each measurement and the active and reactive power consumed by the household.

Second, the Airline Passengers Dataset is a famous and frequently used dataset for time series forecasting. It consists of monthly information from 1949 to 1960 about how many people flew on airplanes during those years. In this particular dataset, passenger counts range from 104 to 622, exhibiting a standard deviation of 119. The mean passenger count is 280. The airline passengers dataset is commonly studied to compare diverse deep learning model for time series analysis. The airline passengers dataset is valuable for several reasons, including its historical significance, which provides a significant amount of data for analysis and forecasting, and its relevance to practical applications. Time series analysis researches frequently use this dataset to assess and compare various models and algorithms for time series forecasting. We picked this known dataset to show that our approach works in other domains.

3.2. Data Understanding and Analysis

Our research objective is to make daily forecasts at a time, which require data at daily interval. Due to the dataset's inherent challenges, such as missing values and high-frequency intervals between measurements, certain steps must be taken to preprocess the data before it can be used for forecasting. Therefore, the dataset was preprocessed to ensure that only relevant data was used for modeling. Data downsampling involves decreasing the time or frequency of a dataset that is beneficial when the original data is collected more frequently than intended for analysis. In our study, we reduced the minutely data to daily intervals to better align with our research goals. This not only simplifies the analysis and modeling processes but also meets our research objectives more effectively. The downsampling process can be summarized as:

Let $X = \{x_0, x_1, \dots, x_n\}$ a time series data with n observations taken at a minute-level frequency, such that x_i represents the data value at time i. Let $T = \{t_0, t_1, \dots, t_n\}$ the corresponding time stamps, where t_i is the time stamp of the ith observation. To downsample the data from minute-level frequency to daily frequency, it can be computed the daily average of the observations for each day shown in equations 1 and 2.

$$\mathbf{X} = \{\mathbf{y_0}, \mathbf{y_1}, \dots, \mathbf{y_n}\} \tag{1}$$

where n is the number of days covered by the time series, and $\boldsymbol{y}_{\mathbf{i}}$ is the average value of the observations for the ith day. The mathematical formula for computing $\boldsymbol{y}_{\mathbf{i}}$ is:

$$\mathbf{y_i} = \frac{\mathbf{1}}{\mathbf{M}} \sum_{\mathbf{j}}^{\mathbf{M}} (\mathbf{x_j}) \tag{2}$$

where M is the number of observations on the ith day, and $\sum_{j}^{M} (x_j)$ represents the sum of the observations on the ith day.

Spline interpolation is widely used interpolation method that is used for imputing missing values in this study. This method works by creating a smooth curve using a set of mathematical expressions called polynomials [21]. The method's application can be particularly advantageous in analyzing time series data, especially in contexts like energy demand prediction. This is because missing values in such data can be attributed to various factors, and a strictly linear trend is not always observed. Spline interpolation is advantageous over other imputation methods because it produces a smooth curve that captures the underlying trend in the data, while also preserving the pattern of the original data. In addition, spline interpolation is less sensitive to outliers than other imputation methods [22] which makes it a more robust approach for handling missing values. The mathematical representation of the cubic spline interpolation formula can be expressed by equations 3-7.

$$\boldsymbol{S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3} \tag{3}$$

$$\boldsymbol{a_i = f_i} \tag{4}$$

$$b_i = \frac{(f_{i+1} - f_i)}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}) \tag{5}$$

$$c_i = \frac{(f_{i+1} - f_i)}{h_i} - \frac{h_i}{3}(2b_i + b_{i+1}) \tag{6}$$

$$d_i = \frac{(b_{i+1} - b_i)}{3h_i} \tag{7}$$

where $S_i(x)$ represents the cubic spline function for the i-th interval, which is defined by the coefficients a_i, b_i, c_i, and d_i. The coefficient a_i represents the value of the function at the i-th data point denoted by f_i. The coefficient b_i and c_i represents the slope and curvature of the function at the i-th data point respectively. The rate of change of curvature is represented by the coefficient d_i.

3.3. Dilated Convolution for Efficient Feature Extraction

LSTM [23] is used as a part of our architecture. LSTM is a type of recurrent neural network (RNN) that is designed to analyze and process sequential relationships in data. LSTM networks can be particularly useful in cases where data has long term dependencies and timing is critical such as natural language processing or speech recognition tasks. The distinguishing feature of LSTMs is their adeptness at discriminative memory or discard previous input values based on the present input and past memory state. This is accomplished through gates that regulate the flow of information in and out of the memory cell. These cells are connected to three gates that have specific functions: input gate, forget gate, and output gate expressed by equations 8-13.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{8}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{9}$$

$$\widetilde{C}_t = tanh\ (W_C \cdot [h_{t-1}, x_t] + b_C) \tag{10}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{11}$$

$$o_t = \sigma(W_0 \cdot [h_{t-1}, x_t] + b_o) \tag{12}$$

$$h_t = o_t * tanh(C_t) \tag{13}$$

where, the forget gate, represented by f_t, determines which information is retained in the cell state, while the input gate, i_t, controls which new input values are added to it. The candidate values for the cell state are denoted by $\widetilde{C}_t$, and the new cell state value, C_t, is a combination of the forget gate and input gate. The output gate, o_t, determines which part of the cell state is activated using the hyperbolic tangent function, and h_t is the output generated from the cell state value and the output gate decision. By utilizing this mechanism, LSTMs can discover and represent intricate patterns and dependencies in sequential data, making them a potent tool for various applications in artificial intelligence and machine learning.

Dilated convolution [24] is a special type of convolution operation that creates gaps between the kernel elements during the convolution process. In this way, a larger receptive field for the network is achieved without an increase in parameters, as shown in Figure 2.

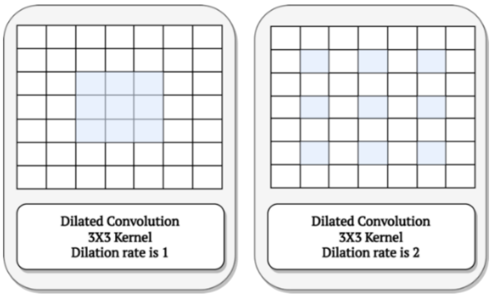

Figure 2. Dilated convolution

In this study, we used dilated convolution for extracting features for LSTM layers. In the context of using dilated convolution with LSTM based neural networks, multiple dilated convolutional layers can be stacked on top of an LSTM layer to extract hierarchical features from time series data shown in Figure 3.

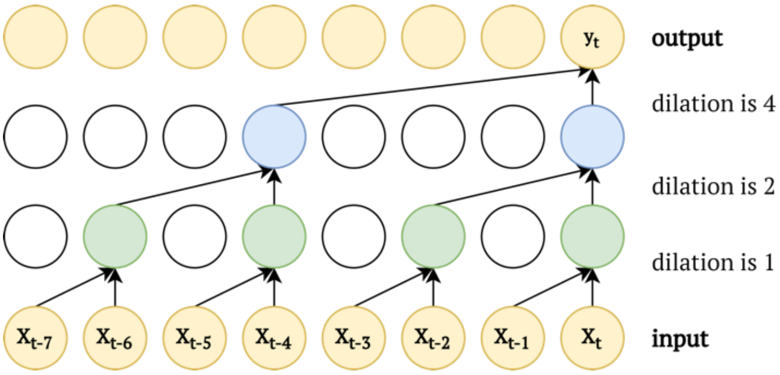

Figure 3. Temporal features

This method can be helpful for understanding connections that persist for a long time in the information, and it can also recognize specific characteristics at various levels. By combining dilated convolutional layers with an LSTM layer, the model can understand intricate patterns and relationships in the data without losing information [25]. It also reduces the risk of overfitting and limiting the number of parameters needed.

3.4. Attention for Preventing Bootleneck

In the analysis of time series data using deep learning models, attention [26] is a technique that lets the model highlight important parts of the data as it processes them. This is achieved by using an additional layer that learns how to weight different parts of the time series based on their relevance. By focusing on the most informative parts of the data, attention mechanisms can improve the model's performance on tasks such as prediction and classification. In this study, we used an attention mechanism to prevent a single connection bottleneck from encoder-decoder LSTM.

Classical Encoder-Decoder LSTM [27] architecture uses a single connection for passing information from encoder to decoder as shown in Figure 4.

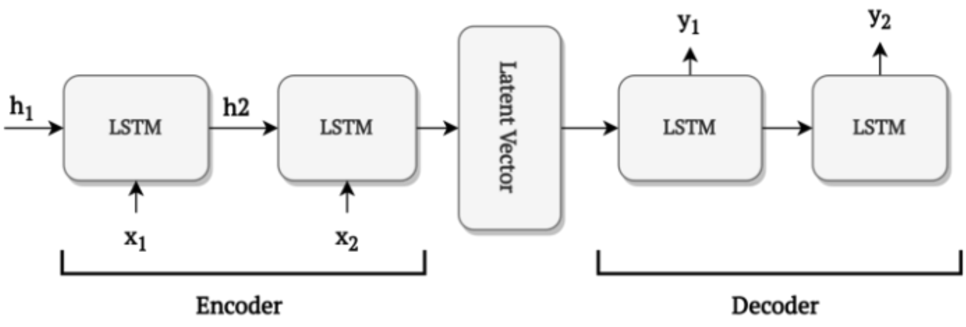

Figure 4. Encoder-decoder LSTM

The utilization of the most common approach with a single connection from the encoder LSTM to the decoder LSTM in a sequence to sequence model can potentially result in a bottleneck for the flow of information between the two network components. This limits the ability of the model to effectively capture the underlying temporal dependencies in the input sequence. To alleviate the issue of information bottleneck caused by a single connection from the encoder LSTM to the decoder LSTM, an attention mechanism was employed for each of the LSTM components which help to improve the information flow by carefully considering and combining the hidden information from the encoder. It does this by giving more importance to the parts that are most relevant to the current decoding step, as illustrated in Figure 5.

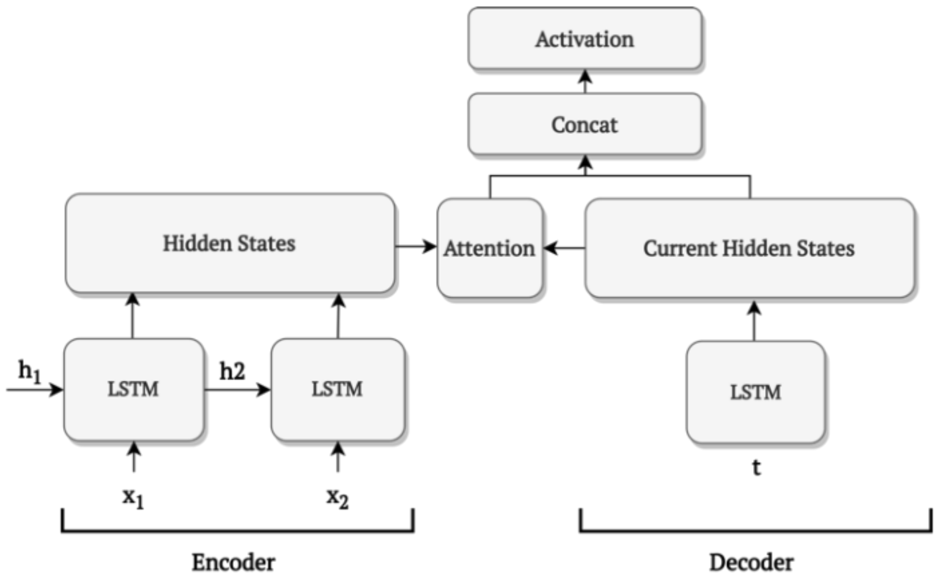

Figure 5. Encoder-decoder LSTM with attention

In time series analysis, the attention mechanism helps deep learning models focus on different parts of the time series data. This may help enhance the modeling of temporal dependencies. The subsequent section of this study will provide a detailed discussion of the experimental results.

4. Experimental Evaluations

In this section, experimental evaluations will be conducted. Specifically, the proposed end-to-end workflow will be evaluated and compared against commonly used deep learning architectures for time series forecasting. This comparative analysis will enable a comprehensive assessment of the effectiveness and efficiency of the proposed approach in addressing the challenges associated with accurate forecasting of both short-term and long-term relationships in time series. The results of this comparison will provide valuable insights into the potential benefits and limitations of the proposed workflow and its ability to outperform existing state-of-the-art techniques in time series forecasting.

4.1. Experimental Setup

Hardware has an important place in training neural networks. The well configured scalable systems helps for efficient training with reducing training time. In this study, the experimental setup consisted of a GIGABYTE GeForce GTX1070 graphics processing unit, 16GB of 3000Mhz DDR4 Dual Kit random access memory, and an INTEL Core i5 8400 2.8GHz 9MB cache 6-core central processing unit. In this study, we used the Adam optimizer, RMSE, and ReLU activation function for our deep learning models. Adam optimizer is a widely employed optimization algorithm in the deep learning approach for updating weights. RMSE is a standard measure for evaluating how well a model performs by calculating the difference between predicted and actual outputs. ReLU is a widely-used activation function in deep learning known for its efficiency and its ability to address the vanishing gradient problem [28].

4.2. Performance Evaluation On Hourly Energy Demand Generation and Weather Dataset

The “Hourly Energy Demand Generation Dataset” is used as a benchmark dataset to evaluate and compare our proposed workflow with other widely used deep learning architectures. To facilitate forecasting of the power consumption for the following day at each time step, the data is transformed into a daily frequency and a window size of 7 is utilized to observe each day in a week. The outcomes of the experiments are detailed in Table 1, which illustrates the performance of the models across varying epochs while maintaining a fixed cell size of 50 to prevent excessive model complexity.

Table 1. Univariate forecasting performance evaluation for hourly energy demand generation and weather dataset

Algorithm	RMSE 100 Epochs	RMSE 200 Epochs	RMSE 300 Epochs
Vanilla LSTM	2837.759	2801.604	2786.482
Stacked LSTM	2731.751	2761.259	2756.587
Bidirectional LSTM	2748.272	2720.271	2723.574
CNN-LSTM	2890.369	2823.410	2809.712
Encoder-Decoder LSTM	2865.107	2840.697	2828.476
Proposed Workflow	**2591.870**	**2596.983**	**2593.754**

Based on the experimental results, our proposed workflow demonstrates a performance improvement ranging roughly from 5% to 11% compared to other deep learning models for the energy demand forecasting task. This improvement in performance can be attributed to the workflow's ability to effectively capture both short-term and long-term dependencies present in the time series data. This feature allows the model to more accurately capture the intricate patterns and relationships that exist within the time series data. Figure 6 visually presents the comparison results, aiding in a better understanding of the obtained outcomes. Also note that, the proposed workflow requires less training time as compared to the others.

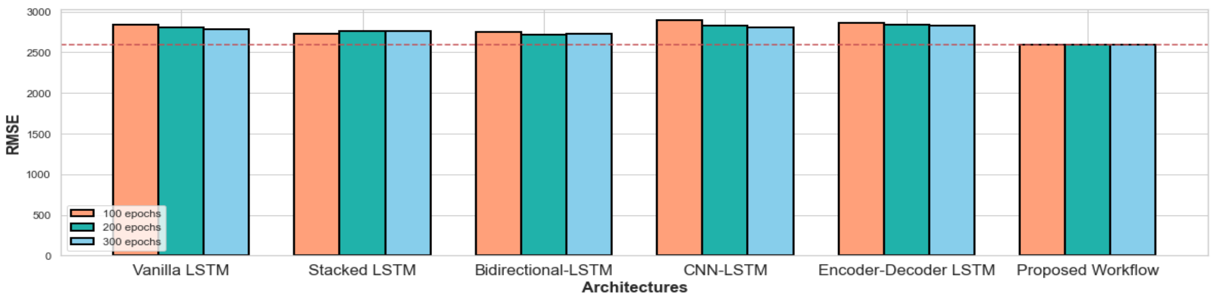

Figure 6. Encoder-decoder LSTM with attention

4.3. Performance Evaluation On Individual Household Electric Power Consumption Dataset

An individual household electric power consumption dataset is used as an another validation dataset. The objective of this task is to forecast power consumption by the household. The data is downsampled into daily frequency and the window size is selected as 7 to observe each day in a week to forecast next day in each time step. The findings are presented in Table 2. It presents the performance of the models for different epochs, while keeping the cell size fixed at 50, to avoid over-complicating the models.

Table 2. Univariate Forecasting Performance Evaluation For Individual Household Electric Power Consumption Dataset

Algorithm	RMSE 100 Epochs	RMSE 200 Epochs	RMSE 300 Epochs
Vanilla LSTM	385.547	397.571	398.474
Stacked LSTM	393.241	401.987	401.889
Bidirectional LSTM	401.517	403.587	402.481
CNN-LSTM	382.674	428.458	432.175
Encoder-Decoder LSTM	375.674	383.487	**387.985**
Proposed Workflow	**366.781**	**378.284**	389.425

Drawing upon the outcomes of the experiments, our proposed workflow demonstrates a performance improvement ranging from 2.4% to 9% compared to other deep learning models for the task of energy demand forecasting. Although the results seem to converge for the encoder-decoder LSTM and the proposed workflow after 300 epochs, the workflow managed to achieve better performance in earlier epochs. That is, the training time is reduced to 1/3, which is an important aspect in training deep learning models. This is because the workflow is designed to capture both the short-term and long-term dependencies of the time series data. It uses an encoder to extract features with dilated convolution from the input time series, the sequential correlation is learned by the LSTM architecture, the most important points of the hidden states are emphasized by the attention layer, and the decoder generates future time steps by mapping to the target time step. This allows the model to effectively capture the complex patterns and relationships in the time series data.

4.4. Performance Evaluation On Airline Passengers Dataset

In order to ensure that the results obtained from a given data analysis workflow can be generalized and applied to other domains, it is important to validate the workflow using widely accepted benchmark datasets. In this study, we used the airline passengers dataset as a benchmark dataset to validate the proposed workflow for time series analysis. The airline passengers dataset is a well-known and widely used benchmark dataset that has been used in many studies for comparing different time series analysis methods. By using this dataset to validate our workflow, we are able to demonstrate that our proposed method is competitive with other architectures that have been previously applied to this dataset. Table 3 displays the quantitative assessment of the performance in terms of root mean squared error based on different numbers of epochs.

Table 3. Univariate forecasting performance evaluation for airline passengers dataset

Algorithm	RMSE 100 Epochs	RMSE 200 Epochs	RMSE 300 Epochs
Vanilla LSTM	42.388	38.891	45.283
Stacked LSTM	44.477	44.941	46.307
Bidirectional LSTM	39.660	33.424	35.964
CNN-LSTM	58.529	51.606	46.432
Encoder-Decoder LSTM	37.819	41.692	44.946
Proposed Workflow	**28.717**	**31.773**	**33.987**

Considering results of our experiments, it appears that the proposed workflow outperforms commonly used deep learning architectures in terms of root mean squared error. The performance improvement is achieved with a range from 27% to 68%. This suggests that the proposed workflow may provide a more effective approach for modeling the underlying structure of the data, particularly in capturing patterns over time and accounting for dependencies between different time steps or observations. In addition, we observed that increasing the epoch value beyond 200 tended to result in overfitting. This highlights the importance of carefully selecting hyperparameters in deep learning models to avoid overfitting and ensure generalizability.

4.5. Overall Performance Evaluation

The preceding sections entail an evaluation of the performance of the proposed attention-based workflow in comparison to other commonly used deep learning architectures for time series forecasting on each dataset separately. The aim of this section is to provide a comprehensive summary of the obtained results, which are exhibited in Figure 7, serving as a benchmark and facilitating a lucid depiction of the achieved outcomes. Benchmarks 1, 2, and 3 denote the performance evaluation of the hourly energy demand generation and weather dataset, individual household electric power consumption dataset, and airline passengers dataset respectively. The RMSE values in Table 1 and Table 2 are relatively high compared to Table 3 due to the larger sensor measurements. Conversely, Table 3 illustrates the deviation of passenger numbers from actual values, which are smaller in magnitude.

There are two major findings. First, our proposed workflow performs better for all three benchmarks. The average gains in performance are 8%, 5.5%, and 42% for benchmarks 1, 2, and 3, respectively. Second, our approach provides results independent of the dataset used. For instance, Bidirectional-LSTM works better on Benchmark 1 than Benchmark 3 while the situation is reversed for Encoder-Decoder-LSTM. While our approach did not exhibit a substantial improvement for some cases, e.g., Proposed Workflow vs. Encoder-Decoder LSTM in Benchmark 3, we showcased its adaptability to different domains, making it scalable and flexible. This underscores its potential utility in diverse industries and domains.

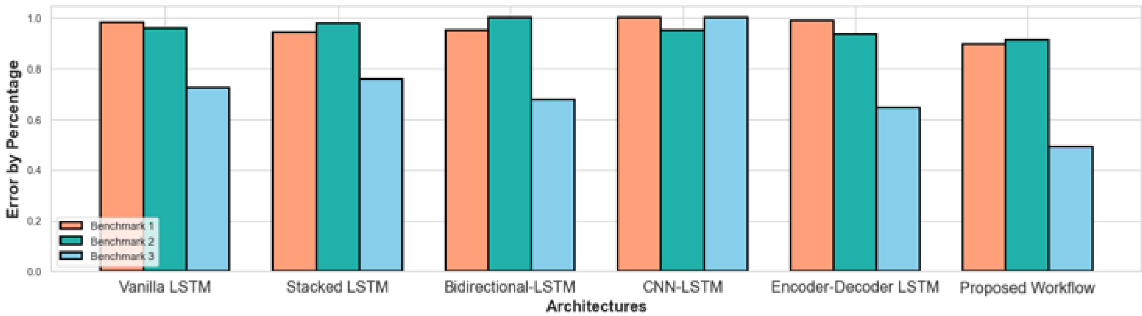

Figure 7. Encoder-decoder LSTM with attention

5. Conclusion

This study highlights the significance of accurate energy demand forecasting in the context of efficient electricity distribution management in smart grids. To this end, we collected and preprocessed a time series dataset of electricity consumption that is primarily associated with smart grid environments. In this research paper, we propose an attention-based time series workflow that integrates dilated convolution and attention mechanisms for time series forecasting in smart grid applications. The workflow employs dilated convolution to extract temporal features from the time series data and attention mechanism to selectively emphasize the significant temporal points in the hidden states. We subsequently evaluated several deep learning-based time series forecasting models and compared them with our proposed workflow. The experimental findings demonstrate that the proposed workflow shows promising results in smart grid domains and can achieve these results in a lower number of epochs, providing a shorter training time. Additionally, we tested our proposed workflow on a dataset from a different domain, i.e., the airline passenger dataset for forecasting customer numbers. The results indicate that the proposed workflow can be adaptable to different domains. Despite several challenges that require future research attention, such as the scarcity of data and the necessity to optimize the model architecture and hyperparameters for varying datasets and forecasting horizons, this study contributes to the burgeoning literature on energy demand forecasting and establishes the groundwork for further research on deep learning-based techniques for smart grid applications.

6. Author Contribution Statement

Author 1 contributed to the writing of the original draft, methodology, visualization and experiments. Author 2, conceptualized the study and was involved in writing, reviewing and editing.

7. Ethics Committee Approval and Conflict of Interest

“There is no conflict of interest with any person/institution in the prepared article”

8. References

[1] V. Kulkarni, S. K. Sahoo, S. B. Thanikanti, S. Velpula, and D. I. Rathod, "Power systems automation, communication, and information technologies for smart grid: A technical aspects review," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 19, no. 3, pp. 1017-1029, 2021.
[2] A. H. Bagdadee, M. Aurangzeb, S. Ali, and L. Zhang, "Energy management for the industrial sector in smart grid system," Ener. Repor, vol. 6, pp. 1432-1442, 2020.
[3] M. S. Hossain, N. A. Madlool, N. A. Rahim, J. Selvaraj, A. K. Pandey, and A. F. Khan, "Role of smart grid in renewable energy: An overview," Ren. and Sust. Ener. Rev., vol. 60, pp. 1168-1184, 2016.
[4] S. Raghavendra, S. Neelakandan, M. Prakash, B. T. Geetha, S. M. R. Asha, and M. K. Roberts, "Artificial humming bird with data science enabled stability prediction model for smart grids," Sust. Comp.: Infor. and Syst., vol. 36, p. 100821, 2022.
[5] D. F. Costa Silva, A. R. Galvão Filho, R. V. Carvalho, F. de Souza L. Ribeiro, and C. J. Coelho, "Water flow forecasting based on river tributaries using long short-term memory ensemble model," Ener., vol. 14, no. 22, pp. 7707, 2021.
[6] Z. Chen, F. Xiao, F. Guo, and J. Yan, "Interpretable machine learning for building energy management: A state-of-the-art review," Adv. in Appl. Ener., vol. 100123, 2023.
[7] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," Comp. Scie. Rev., vol. 40, p. 100379, 2021.
[8] Y. Xu and L. Zhao, "Inception-LSTM Human Motion Recognition with Channel Attention Mechanism," Comput. and Math. Meth. in Med., 2022.
[9] J. Shi and Z. Wang, "A hybrid forecast model for household electric power by fusing landmark-based spectral clustering and deep learning," Sust., vol. 14, no. 15, pp. 9255, 2022.
[10] M. M. Keskin, F. Irım, O. Karaahmetoğlu, and E. Kaya, "Time series prediction with hierarchical recurrent model," Sig., Ima. and Vid. Proce., vol. 17, no. 5, pp. 2121-2127, 2023.
[11] I. S. Oh and J. S. Lee, "Dense sampling of time series for forecasting," IEEE Acc., vol. 10, pp. 75571-75580, 2022.
[12] G. Hebrail and A. Berard, "UCI machine learning repository: Individual household electric power consumption dataset," EDF R&D, vol. 30, no. 08, 2012.
[13] G. Garnero and D. Godone, "Comparisons between different interpolation techniques," The International Archives of the Photogr., Remote Sens. and Spatial Infor. Scien., vol. 40, pp. 139-144, 2014.
[14] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," Comp. Stat. & Data Analy., vol. 54, no. 4, pp. 1167-1178, 2010.
[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comp., vol. 9, no. 8, pp. 1735-1780, 1997.

[16] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.
[17] R. Xi, M. Hou, M. Fu, H. Qu, and D. Liu, "Deep dilated convolution on multimodality time series for human activity recognition," in 2018 International Joint Conference on Neural Networks (IJCNN, pp. 1-8), July 2018.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," in Adv. in Neural Infor. Proce. Syst., vol. 30, 2017.
[19] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Adv. in Neural Infor. Proce. Syst., vol. 27, 2014.
[20] M. S. Ko, K. Lee, J. K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting," IEEE Transac. on Sustain. Ener., vol. 12, no. 2, pp. 1321-1335, 2020.
[21] A. V. Seliverstova, D. A. Pavlova, S. A. Tonoyan, and Y. E. Gapanyuk, "The time series forecasting of the company's electric power consumption," in Advances in Neural Computation, Machine Learning, and Cognitive Research II: Selected Papers from the XX International Conference on Neuroinformatics, October 8-12, 2018, Moscow, Russia, pp. 210-215, Springer International Publishing, 2019.
[22] A. Shobol, M. H. Ali, M. Wadi, and M. R. TüR, ‘‘Overview of big data in smart grid,’’ in Proc. 8th Intl. Conf. Renew. Energy Res. Appl. (ICRERA, pp. 1022–1025), 2019.
[23] D. C. Sekhar, P. R. Rao, R. Kiranmayi, "Large Scale Predictive Analytics based Real-Time Energy Management and Enhance Power Quality in Smart Grid," in Proceedings of the 2022 IEEE 7th International Conference for Convergence in Technology (I2CT), Apr. 1-5, pp. 1-5, 2022.
[24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," arXiv preprint arXiv:1511.07122, 2015.
[25] R. Xi, M. Hou, M. Fu, H. Qu, and D. Liu, "Deep dilated convolution on multimodality time series for human activity recognition," in 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-8, July 2018.
[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," in Adv. in Neural Inform. Proce. Syst., vol. 30, 2017.
[27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Adv. in Neural Infor. Proce. Syst., vol. 27, 2014.
[28] M. S. Ko, K. Lee, J. K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep concatenated residual network with bidirectional LSTM for one-hour-ahead wind power forecasting," IEEE Transac.on Sustain. Ener., vol. 12, no. 2, pp. 1321-1335, 2020.