Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5, 1014-1037
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i5.1803

Univariate vs multivariate predictability analysis of electricity consumption by LSTM approach: An empirical evaluation

Fatma Yaprakdal
Information Systems Engineering Department, Mehmet Akif Ersoy University, Türkiye; fatmayaprakdal@mehmetakif.edu.tr (F.Y.)

Abstract: Precise prediction of EC is crucial for planning, managing, and cost-effective operation of power grids, as it is a time series problem. In recent years, numerous studies have analyzed the behavior and quality of time series forecasts in different areas, including the EC. New models or variants of traditional algorithms have also been proposed, usually taking advantage of the increasing amount of data available and of today's computing power, such as LSTMs. One of the most important characteristics in the selection of a forecasting method is the number of variables that must be taken into account in the prediction of the time series since most of these variables are subject to external influences. EC is dependent on a number of external factors, such as climatic factors, economic factors, and the spot price of electricity. The EC dataset may be either univariate or multivariate. When dealing with univariate time series data, it's important to utilize specific methods that take only the historical values of the variable into consideration for accurately estimating its pattern. Prediction methods that analyze dependencies and correlations between variables to predict future values are also suitable for multivariate time series. Nevertheless, these approaches usually need more time to compute and train, and they might not even be the most appropriate way to go, because the increased complexity of the model used could outweigh any possible improvement in prediction accuracy. In this paper, a thorough comparison study was conducted to analyze the effectiveness of univariate and multivariate predictive analysis on two separate sets of EC data at hourly and daily intervals. To accomplish this, the LSTM algorithm was utilized, which has recently been widely used and recognized as the best-performing algorithm in EC forecasting studies. The comparative analysis is complemented by a comprehensive literature review, meticulously presented in a tabular format, to offer a comprehensive understanding of the univariate and multivariate forecasting methodologies and their respective outcomes. This study stands out due to its incorporation of an extensive literature review to support the experimental research, ensuring a thorough evaluation. Based on experimental studies, univariate forecasting analysis outperformed multivariate forecasting analysis for both hourly and daily interval data sets. Furthermore, the R-squared results of the univariate and multivariate predictive analyses conducted with the hourly data set are significantly higher than those of the same predictive analyses in the daily interval.

Keywords: *Electricity consumption, LSTM, Time-series forecasting, Univariate, Multivariate.*

1. Introduction

Electricity demand forecasting, which helps to balance the future electricity needs of different electricity-consuming sectors, is a critical element in the planning, management, and development of the electricity sector. As the power market becomes increasingly deregulated, it is more important than ever for utilities to produce more accurate EC predictions [1]. In competing energy sectors, EC is characterised by linear or non-linear long-term tendencies, different periodicity, non-stationary mean

* Correspondence: fatmayaprakdal@mehmetakif.edu.tr

and variance, peaks or troughs (extremes), high volatility and calendar effects. The literature contains a large number of models, among which are statistical, ML, econometric and hybrid models, designed to forecast EC. These models differ in their methodology, complexity, and performance. The comprehensive list of linear (AR, ARMA, ARIMA, SARIMA, and ARIMAX) and non-linear time series models (ARCH, GARCH, SGARCH, TGARCH, and EGARCH) for electricity demand forecasting provides a wide range of options to choose from in conducting forecasting. It's constructive to include parametric and non-parametric regression-type models, as well as models based on exponential smoothing, as they offer diverse approaches for addressing specific requirements in electricity demand forecasting [2]. While neural networks have many advantages over these conventional time series models for dealing with nonlinear and nonnormally distributed data, which are often encountered in real problems, their disadvantage is that they assume that all of the inputs and outputs are independent, even for continuous data. The assumption being made neglects the predictive potential offered by the dependency relationship between energy consumption and sequential data. It's essential to consider the role of DL architectures, which are adept at tackling the specific challenges encountered in LF, including non-linearity, periodicity, and seasonality, as well as the sequential dependency between sequences of consumption data. LSTM networks, a variant of DL introduced by Hochreiter and Schmidhuber, are specifically engineered to grasp the long-term dependencies inherent in sequential data. By leveraging internal memory to store long-term dependencies, LSTMs are well-suited for addressing problems characterized by sequence-dependent behavior, such as electricity demand forecasting [3].

When selecting a forecasting method, it's crucial to consider the variable numbers that will be involved in the time series prediction, given that most variables are influenced by external factors. For univariate series, there are specific methods that consider only the past values of the variable itself to estimate its evolution. Other forecasting algorithms analyze dependencies and interactions with other variables to predict future values and are more suitable for multivariate time series. Since most variables can be influenced by external factors, these are the most commonly used in real-world applications. However, these methods usually require longer computation and training times, and may not always be the best solution because the improvement in prediction accuracy is not offset by increased model complexity [4]. On the other hand, given the lack of multivariate data (such as temperature) in many practical datasets, we need better univariate prediction models for time series LF [5]. Although models can differ dramatically among univariate and multivariate systems, most ML and DL models can handle both indistinctly [6].

In this paper, we perform a detailed comparative study to evaluate how effective both univariate and multivariate predictive analysis methods are on two different EC datasets. We examine the impact of these methodologies at both hourly and daily intervals, aiming to provide a richer understanding of their effectiveness in predicting EC patterns. In order to accomplish our goal, we applied the LSTM algorithm. This algorithm has garnered significant recognition due to its effectiveness in various studies focusing on EC forecasting within the existing literature. The comparative analysis is enhanced by a thorough literature review, meticulously presented in a tabular format, to provide a comprehensive understanding of univariate and multivariate forecasting methodologies and their respective outcomes. On the contrary, due to the likelihood of the pattern and behavior of EC recurring in the future, forecasts for EC are established upon past observations. Consequently, when conducting time series forecasting, it is crucial to meticulously select the most appropriate past observations that can effectively serve as predictors of expected future values. This process involves identifying historical data points or patterns that have the most significant impact on forecasting future values and ensuring that they are accurately incorporated into the predictive model. In this study, we determined the lag periods by carefully examining the time intervals established in previous relevant research studies. We specifically considered the data characteristics studied in the literature to ensure that the chosen lag periods were appropriate for our analysis [3]. The length of the prediction is crucial for its accuracy. It determines how many future data points the forecast should cover. If the predicted length is zero, the autoregressive model focuses on forecasting just one step ahead. However, a predicted length greater than zero indicates a multi-step forecasting problem. This paper uses one-step-ahead forecasting for clarity [7].

The structure of the sections of the paper is as follows: first, a relevant and necessary technical details for time series electricity forecasting is presented. Section 3 presents a review of LSTM approaches for univariate and multivariate electricity time series forecasting. Section 4 provides details on the experimental study and presents the plots and experimental results concerning a common performance measure, coefficient of determination (R-square), obtained from this study, along with comments and discussions. Finally, the last section will focus on drawing conclusions and outlining future work.

2.Technical Details

2.1. Forecasting of Time Series

Time series consists of a sequence of data points, typically measurements taken at successive points within a time period. These data points are usually taken at regular intervals, and the order of the points is crucial. Time series forecasting is a predictive modeling technique that consists of analyzing time series historical data in order to make informed predictions about future values. This analytical approach is indispensable for informed decision-making across a wide range of fields, including finance, weather forecasting, and resource planning. By identifying patterns and trends within the historical data, time series forecasting allows for the anticipation of future outcomes, enabling businesses and organizations to make strategic and proactive decisions. From predicting stock prices and sales trends to anticipating weather patterns and demand for resources, the application of time series forecasting plays a vital role in optimizing planning and resource allocation. Despite time being a continuous variable, in a time series, the values are discrete and sampled at fixed intervals, enabling analysis and forecasting based on these specific data points [7].

Time series models can handle either single-variable (involving only one time-dependent variable, like temperature over time) or multi-variable (involving multiple time-dependent variables, such as temperature, humidity, and air pressure over time) data. Most ML and DL models are capable of working with both types of time series data. However, it's important to note that the models can vary significantly in their implementation and performance depending on whether they are applied to a univariate or multivariate system.

2.1.1. Forecasting of Univariate Time Series

Forecasting univariate time series creates extrapolations for a single variable based on past time series observations of the same variable. Despite the geometric increase in data availability, univariate forecasting is still the basis for decision-making in many organizations. Improving the performance of such forecasts is critical to reducing operational, tactical, and strategic planning costs [9]. Univariate modeling's primary advantage is that it predicts future events based on historical data points and how they have behaved [10].

Given y = y(t-L), ..., y(t-1), y(t), y(t+1), ..., y(t+h) is a univariate series of 'L' the past data values, where every y(t-i), for i = 0, ..., L, indicates the stored values of the variable y for the time t - i. Forecasting consists in predicting the future value of the y(t+1), indicated by $\hat{y}$(t+1), in order to minimize the error, which is usually given as a function of y(t+1)-$\hat{y}$(t+1).

This forecast is also possible if the horizon of the forecast, h, is greater than 1, i.e. if the aim is to forecast the h subsequent values after y(t), i.e. y(t+i), with i = 1, ..., h. When the following function is minimized, the best forecasting is achieved here [6].

$$\sum_{i=1}^{h}(y(t+i)-\hat{y}(t+i))$$

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

2.1.2. Forecasting of Multivariate Time Series

Multivariate time series forecasting has wide applications such as traffic flow prediction, electricity demand forecasting, stock market forecasting, etc., and a large number of forecasting models have been developed [11]. Multivariate time series forecasting models have transitioned from classical statistical methods to DL methods over the past several decades [12]. The Multivariate time series forecasting problem is very common in real-life applications. Intuitively, it is more complex since it has two or more variables [13]. Multivariate time series may be expressed in matrix form in the following way,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ y_T \end{pmatrix} = \begin{pmatrix} y_1(t-L) & \cdots & \cdots & y_1(t-1) & y_1(t) & y_1(t+1) & \cdots & \cdots & y_1(t+h) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n(t-L) & \cdots & \cdots & y_n(t-1) & y_n(t) & y_n(t+1) & \cdots & \cdots & y_n(t+h) \end{pmatrix}$$

where yi(t-m) shows time series set, with i = 1, 2, 3, …, n being m = 0, 1, 2, …, L the past data and current sample, and m = -1, -2, ... , -h the future value of h. In time series analysis, it is customary to designate one series as the target time series and the others as independent time series [6].

2.2. LSTM Approach

Sequential data, such as word sequences in machine translation, audio data in speech recognition, or time series in forecasting, all exhibit a common characteristic: they possess a temporal dependency. Traditional FFNNs are unable to accommodate these dependencies, a problem that RNNs are specifically designed to address. RNNs handle this issue by incorporating both past and current data in their architecture. Furthermore, in time series methods, the data must be analyzed to determine whether it is stationary or not. In contrast, LSTM can give good results regardless of whether the data is stationary or not. The configuration of data inputs and outputs in a network can be classified into different categories based on the relationships between them. These categories include one-to-one (where there is one input and one output), one-to-many (where there is one input and multiple outputs), many-to-one (where there are multiple inputs and one output), and many-to-many (where there are multiple inputs and multiple outputs). These classifications are important for understanding how information is processed and transmitted within the network.

Standard basic RNNs suffer from the vanishing gradient problem, where the gradient decreases with the increasing quantity of layers. When neural networks have a large number of layers, the gradient becomes very small, which stops the network from learning effectively. As a result, these networks have short-term memory and struggle with learning from long sequences that require remembering all the information in the sequence. To address this issue, LSTM recurrent networks have been developed. LSTMs use three gates to retain important information for a longer time and filter out irrelevant information. These gates are the forgetting gate (ft), the updating gate (it), and the output gate (ot). The ft decides which of the information needs to be excluded. or stored. Any value near zero indicates that past information has been forgotten, while a value that is very close to 1 means that it will be kept. The it decides which new information the $\widetilde{c_t}$ should be used for updating the c_t memory state. Thus, the ct is updated using both ft and it. The final step involves the ot determining the input for the next hidden unit. Finally, the ot decides which output will be the next hidden unit's input.

After passing through the sigmoid activation function (σ), the current input and xt values are used to compute all the gate values. The tanh activation function is then applied to calculate the $\widetilde{c}_t$, which is used for updating. The equations defining an LSTM unit are:

$$\widetilde{c_t} = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
$$c_t = (i_t \times \widetilde{c_t}) + (f_t \times c_{t-1})$$

ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

$$h_t = o_t \times \tanh(c_t)$$

where Wi, Wf, and Wo represent the weights that drive the behavior of the input gate, forget gate, and output gate respectively. Similarly, bi, bf, and bo are the biases corresponding to the input gate, forget gate, and output gate. Additionally, Wc and bc denote the weight and bias of the memory cell candidate ($\widetilde{c_t}$).

3. The Literature Survey

This part of the paper gives an overview of the literature on univariate and multivariate predictive analysis studies that estimate the EC using LSTM-based approaches. As part of the literature review, the Google Scholar database was searched using the keywords 'Univariate time series electricity load demand consumption forecasting LSTM' and 'Multivariate time series electricity load demand consumption forecasting LSTM'. From the publications obtained as a result of the search, we review recent studies and the most prominent studies in indexed journals. The literature outlines various input parameters used for electricity forecasting, resulting in the development of complex models with an expanded number of variables. Additionally, particular studies focus on utilizing a single input variable for prediction of future values on the basis of its actual values [15].

In this context, three separate tables have been created: Table 1 presents multivariate time-series models, the second table presents univariate time-series models, and the last table presents studies that include both of these topics. The tables include the forecast models used in the study, a summary of the study, the lagged time steps used in the forecast models, how many steps ahead the forecasts are made, and the results obtained from the study. All the papers mentioned in the table have been analyzed in detail and the information not mentioned in these papers is indicated as 'unspecified' in the table. Various forecasting designs and methods for multivariate and univariate time series EC forecasting using LSTM-based approaches have been introduced in numerous studies. Looking at Table 1, EC forecasts as multivariate time-series models are mostly short-term forecasts, but at the residential level. In these studies, the simple LSTM model was used in comparison with other ML-based models or statistical models. Apart from this model, LSTM models such as variations of the LSTM model (Bi-LSTM, LSTM encoder-decoder, etc.), deep LSTM, and hybrid or ensemble models created by combining the LSTM model with other ML-based models have been applied. In the residential level studies, data sets of a single building or multiple buildings are studied and the correlation of external factors with EC is analyzed. Although the results vary according to the dynamics of each data set, the results of these forecasting studies with LSTM-based approaches are quite successful. Looking at Table 2, EC forecasts as univariate time-series models are realized in short and medium-term periods, but at the aggregated system level. In these studies, as in multivariate time-series models, the simple LSTM model was used in comparison with other ML-based models or statistical models, and as LSTM models other than this model, LSTM model variations (Bi-LSTM, LSTM encoder-decoder, etc.), deep LSTM and hybrid or ensemble models created by combining the LSTM model with other ML-based models were applied. Table 3 shows the papers in which both multivariate and univariate time-series models were studied together. In some of these papers, it is stated that LSTM-based forecasting models are more successful on multivariate models. In multivariate models, it has been observed that more successful results are obtained with features that are determined by feature selection and have a high correlation with EC forecast.

Table 1.
Multivariate time-series EC forecasting papers.

Ref.	Model	Summarize	Lagging steps	Forecasting horizon	Results
[16]	A single scalable LSTM	A single but complex LSTM model was proposed, capturing key features of individual consumption and household information.	24 hrs	24 hrs	The model achieves promising results against competitive benchmarks, outperforming them on average by more than 20% over all test periods and all test measures in their back-testing experiment.
[17]	ARIMA, SARIMA and LSTM	Analyzing the collected smart meter data to predict the daily EC using ARIMA, SARIMA, and LSTM provides insights into factors influencing EC to support decision-making.	lags up to 7 days, for hrs 12 lags distributed in quarter	1 step	The results indicate a strong positive correlation between EC and humidity, and a significant negative correlation between EC and temperature. Dew point and UV index, as well as cloud cover and visibility index, show multicollinearity with temperature and humidity respectively. Overall, LSTM outperforms ARIMA and SARIMA with an average MAE of 0.23.
[18]	CNN-LSTM	The performance using a CNN-LSTM model to predict residential EC was benchmarked against LSTM, GRU, Bi-LSTM, and Attention LSTM models.	60 min.	60 min.	The CNN-LSTM model achieved the highest performance with the MSE of 0.37, outperforming LSTM, GRU, Bi-LSTM, and Attention LSTM across minute, hour, day, and week unit resolutions.
[19]	A CNN-LSTM model	Developed a CNN-LSTM model using multivariate augmentation. Conducted state-level analysis and training to demonstrate forecasting accuracy for regional energy consumption.	30 days	1 day	The pooling layer of the 1D CNN reduced noise, lowering RMSE and MAPE values. The LSTM layer, receiving inputs for each time step, was ideal for processing time series data. Extensive experiments and ablation studies were conducted to showcase the advantages of the proposed CNN-LSTM architecture coupled with multivariate augmentation for provincial EC time series forecasting.
[20]	DLSTM	For big data, a DL-based price and demand forecasting model was proposed. The suggested DLSTM was benchmarked against traditional ANN models:: NARX Variables and ELM.	unspecified	unspecified	The DLSTM network was trained with monthly data and outperformed other methods in accuracy. The DLSTM has an MAE of 2.9, while the benchmark method has an MAE of 9.7. The proposed method's NRMSE is 0.087, compared to

					the benchmark's MAPE of 0.2.
[21]	LSTM	A time series forecasting model utilizing LSTM with social and weather-related variables was proposed. The performance of the LSTM model was benchmarked against the SVR, ANN, ARIMA and MLR models.	24 months	24 months	The LSTM model outperformed four benchmark models in MAPE.
[22]	CEEMDAN-SE and LSTM	This study developed a model called CEEMDAN-SE-LSTM to forecast ultra-short-term electricity load in Changsha, China. The model takes into account meteorological and holiday factors.	2, 4, 6, 8, 10, 12, 24, 36, 48, 60, and 72 hrs	4 & 8 hrs	The CEEMDAN-SE-LSTM model demonstrated superior performance compared to the ARMA, LSTM single-prediction, EEMD-LSTM, and CEEMDAN-LSTM models, with significantly lower RMSE, MAE, and MAPE values of 62.102, 47.490, and 1.649% respectively.
[23]	Bi-LSTM-based encoder-decoder with an attention mechanism	A DL structure using Bi-LSTM layers and a temporal attention mechanism is proposed to learn long-term dependencies and hidden correlation features in multivariate temporal data.	Arbitrary lengths	from 1 to 6 steps	The experiment results from five multivariate time series datasets provided compelling evidence that the proposed model has the capability to accurately forecast multi-step time series values, irrespective of short-term or long-term time step conditions.
[24]	SVR, LSTM, GRU, CNN-LSTM, CNN-GRU	SVR, LSTM, GRU, CNN-LSTM, and CNN-GRU models were compared to predict energy consumption data of smart homes.	1 step	1 step	As the data amount increases, SVR's performance degrades more than DL techniques, indicating that ML techniques may not be suitable for the task. The CNN-GRU architecture performed best for daily granularity, while the LSTM was best for hourly granularity. The LSTM outperformed the CNN-GRU architecture by 0.4% in terms of MAE, while the CNN-GRU outperformed the LSTM by 17.4% for daily granularity.
[25]	LSTM encoder–decoder	A day-ahead forecasting technique for individual residential load demand based on the LSTM encoder-decoder architecture with both past and future exogenous inputs was presented.	48 steps (half an hour)	48 steps (half an hour)	The proposed model outperformed three selected benchmark methods NARX, ANN, Block LSTM, and Naïve Seasonal by reducing the mean absolute scaled error by up to 8%.
[26]	LSTM, Bi-LSTM and GRU	LSTM, Bi-LSTM, and GRU were compared for the prediction of Miscellaneous Electric Loads (MEL).	unspecified	1 day and 1 week	The study's results indicated that both the Bi-LSTM and GRU models outperformed the LSTM model, especially for longer prediction horizons.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

[27]	CNN-LSTM	An STLF integrated the LSTM and CNN models, taking into account each's advantages. The suggested CNN-LSTM approach was benchmarked against the LSTM, RBFN, and XGBoost models.	7 days	24 hrs, 48 hrs, 1 week, 1 month	The CNN-LSTM approach outperformed the LSTM, RBFN, and XGBoost models in terms of MAE, RMSE, and MAPE values. It is observed that the CNN-LSTM approach effectively handles long-sequence EC data and predicts future EC for a substantial period of time.
[28]	Parallel DLSTM- CNN	The LSTM-CNN combination has enabled the development of an advanced approach for short-term load forecasting, known as PLCNet, which has promising potential for accurately predicting load dynamics.	Malaysian case: 72 hrs, 4 days, 10 days. German case: 7 days, 10 days, 30 days.	Next 24 and 48 hrs, and next 10 days (Malaysian case). Next 7, 10, and 30 days (German case).	The PLCNet model demonstrated notable improvements in accuracy across different time horizons, boosting performance from 82.49% to 91.31% for the German data and from 94.16% to 98.14% for the Malaysian data.
[29]	LSTM	The LSTM model was used to predict residential EC by aggregating loads from an optimized selection of households by the OPTICS algorithm. The proposed method was compared with SVR-based and BPNN-based methods.	6 steps (3 hrs) and 48 steps (24 hrs)	half an hour	The results showed that the proposed method performed the best in terms of the MAPE metric among the benchmarking methods in all cases. Moreover, the resulting MAPE value is less than 10%, which is quite sufficient for microgrid dispatch.
[30]	LSTM network with transfer learning	The authors utilized the LSTM network with XCORR transfer learning for LF. They cross-correlated the time series to determine the training data set and used transfer learning to predict a new dataset.	1 step (15-minute)	1 step (15-minute)	The RMSE, MAPE and MAE scores indicated that, compared to the RF, XGB and LGBM models, the LSTM with transfer learning was successful in predicting EC in buildings with the least amount of energy data..
[31]	LSTM, RF, and XGBoost	LSTM, RF and XGBoost were investigated to predict energy consumption in Korea. These methods were employed on a time series prior to and following the COVID-19 pandemic.	unspecified	unspecified	The results indicated that the LSTM model had lower RMSE and MAPE in the first period, while RF had lower RMSE and MAPE in the second period. In summary, LSTM performed best in the first period while RF performed best in the second period.
[32]	DNN, Bi-GRU-	This study explored the STLF of five	1hr, and	1 hr	The test results indicate achieving a MAPE of

	FCL, GRU-FCL, Bi-LSTM-FCL, and CNN	aggregation levels (3, 10, 30, 100, and 479) of a dataset of 479 residential dwellings. The sample sizes per level were 159, 47, 15, 4, and 1. Five DL approaches were used for each aggregation level.	24, 48, 72 and 96 hrs		2.47-3.31% close to the country level at the highest aggregation and maintaining less than 10% at 30 aggregated dwellings. The DNN showed the highest performance, followed by the Bi-GRU-FCL with nearly 15% faster training time and 40% fewer learnable parameters.
[33]	CNN and Bi-LSTM	The proposed model called EECP-CBL combines CNN and Bi-LSTM to predict EC. Two CNNs extract information from household variables, followed by a Bi-LSTM module making predictions.	Un-specified	Un-specified	The EECP-CBL framework outperformed LR and LSTM approaches in predicting EC across various time periods and performance metrics.

Table 2.
Univariate time-series EC forecasting papers.

Ref.	Model	Summarize	Lagging steps	Forecasting horizon	Results
[34]	ML and LSTM-based neural network	ML- and LSTM-based approaches in different formations were used to develop predictive methods for short to medium term LF.	From 1 to 99 steps	For the short term: from a couple of days to 2 weeks, For the medium term: a couple of weeks to a couple of months	The use of only optimised time-lagged features in an LSTM model effectively captured the intricacies of the complex time series and resulted in reduced MAE and RMSE in medium to long-term prediction for a large metropolitan area.
[35]	LSTM, GRU, TCN	An innovative approach using ensemble learning has been implemented to forecast monthly EC. This method combines three successful models in the field: LSTM, GRU, and TCN.	12 months	1 month	The proposed ensemble learning models significantly improved the prediction performance compared to the different models separately in terms of MAE and MAPE. However, TCN obtained the best results in terms of RMSE due to the low variance of the errors.
[36]	ARIMA -LSTM, ARIMA -GRU	Peak EC was forecasted using a hybrid approach that combined traditional time series forecasting (ARIMA) with DL methods (LSTM, GRU). ARIMA modeled the	24 months	1 month	The hybrid approach, ARIMA-LSTM, gave the most accurate predictions with an RMSE value of 7.35, outperforming the hybrid ARIMA-GRU approach with an RMSE value of 9.60. In comparison, the individual models (GRU, LSTM and ARIMA) obtained higher RMSE values (18.11, 18.74 and 49.90 respectively).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

		trend, while LSTM and GRU interpreted fluctuations. The combined outputs provided the final prediction.			This highlights the superior performance of the hybrid approaches over the single approaches.
[37]	Seq2Seq LSTM	A time-series clustering approach using a multi-stage Seq2Seq LSTM LF strategy for households was proposed to improve the efficiency of the demand response programme.	from 1 to 200 steps	1 step	The proposed model performed best when clustering was combined with Seq2Seq EC prediction at 60, 120 and 180 steps, outperforming the other models.
[38]	LSTM, Bi-LSTM	The comparison between the LSTM model and the Bi-LSTM for a univariate time series STLF model involved the utilization of four different datasets with varying contexts and scales to comprehensively assess the robustness of the models.	1 step	1 step	BLSTM performed better than LSTM models in predicting time series EC, despite requiring a longer training time for better results.
[39]	LSTM	An LSTM neural network is proposed for STLF. Optimal hyperparameter values were obtained using random search and the Coronavirus Optimization Algorithm. The LSTM model was then used to predict electricity demand for a 4-hour forecast horizon.	168 steps (10-minute)	24 steps (10-minute)	The results show highly accurate predictions, with a MAPE of less than 1.5. The smallest errors were found when comparing different models, including LR, DT, ensembles of trees, and two DNNs: a D-FFNN with optimized random search and a temporal fusion transformer with optimized sampling algorithm.
[40]	LSTM, GRN, and CNN	LSTM, GRN and CNN models were benchmarked to forecast 1-7 days ahead of daily EC.	7, 14, 21, 28, 182, 364, 546, 728 days	1 to 7 days	The LSTM algorithm achieved the best forecasting performance. In the test set results, it had an R-squared of 0.94 for one-day forecasting, dropping to 0.73 for seven-day forecasting.
[41]	LSTM	An LSTM-based LF framework was proposed for	2, 6, 12 steps	unspecified	The suggested LSTM model outperformed the other benchmarking models in the STLF problem for

		individual household LF, and its performance was compared with various benchmarks, which include the state of the art in LF.	(half an hour)		individual households by an average MAPE of 44.06% and for aggregated households by an average MAPE of 8.58%.
[3]	Multi-Sequence LSTM-RNN tuned by the MSA	Metaheuristic search-based algorithms were utilized to optimize the tunable LSTM hyperparameters for EC. GA and PSO were used to optimize hyperparameters for predicting EC in big data applications.	From 1 step to 2880 steps (half an hour)	unspecified	Upon conducting statistical analysis of the results, it was revealed that the multi-sequence DL model, fine-tuned by metaheuristic search algorithms, yielded significantly more accurate results compared to the benchmark ML models (SVR, RF, and ANN) and the manually configured LSTM..
[42]	An ensemble LSTM combined with SWT	A hybrid DL model was created by combining an ensemble LSTM model with the SWT approach. Multiple SWTs made the original EC data stationary, and LSTM was applied to produce prediction results.	5, 10, 20, and 30 minutes step sizes	unspecified	Compared to other advanced methods, including the persistent method, SVR, LSTM neural network, and CNN-LSTM, the proposed method demonstrated superior performance across three error metrics (RMSE, MAPE, and MBE).
[43]	Central Energy Authority trend-based model, SARIMA, LSTM, Facebook Prophet	Four time-series models were compared to predict total and peak monthly EC in India: the existing trend-based model of India's Central Energy Authority, SARIMA, LSTM, and Facebook Prophet.	108 months	24 months	LSTM predictions are found to be unreliable due to large prediction errors, and predictions tend to become highly unstable after 24-time steps. Facebook Prophet predictions performed best in reproducing the temporal features present in the observed data. Therefore, it is suggested that the Facebook Prophet model should be preferred over others in demand forecasting.
[44]	ARIMA	ARIMA and CNN-Bi-LSTM	unspecified	1 month	The results indicated that only the CNN-LSTM

	and CNN-Bi-LSTM	methods were proposed to predict medium-term EC. Hyperparameters were tuned for ARIMA and neural network models to enhance model accuracy.			combination did not perform well. On the other hand, the ARIMA model produced accurate results and can be used in real-world scenarios where data patterns remain relatively consistent. Additionally, the combination of CNN and Bi-LSTM yielded favorable results with lower MSE and RMSE for medium-term EC prediction following the implementation of ARIMA.
[45]	ICMD-ANN-Encoder-Decoder-based LSTM hybrid model	A hybrid multi-algorithm framework is developed by incorporating ANN, Encoder-Decoder Based LSTM, and ICMD. This model was compared with single models (ANN, RFR, LSTM), hybrid models, and three decomposition-based hybrid models.	From 1 to 15 days	1 day	According to statistical score metrics, the hybrid ICMD-ANN-EDLSTM model performed better than other benchmark models. Furthermore, the results showed that the hybrid ICMD-ANN- Encoder-Decoder Based LSTM model can not only detect seasonality in electricity demand data but also predict and analyze electricity market demand.

Table 3.
Multivariate time-series EC forecasting papers.

ef.	Model	Summarize	Lagging steps	Forecasting horizon	Results
[46]	DNN, RNN, CNN, and LSTM models with forecasting error correction	This study proposed the use of DL techniques to predict EC in two industrial buildings that have varying usage profiles. It analyzed the correlation between EC and meteorological data per season for the related buildings.	1 hr	24 hrs	DNN, RNN, and LSTM outperformed by utilizing two years of EC and weather data instead of one year of EC.
[47]	XGBoost, LSTM, and SARIMA	Three essential techniques (univariate, multivariate, and multistep) were investigated to predict EC using three leading methods: XGBoost, LSTM, and	1 step	1 step for multivariate and 72 steps for univariate	XGBoost performed well for univariate and multistep predictions compared to ARIMA variations, which performed better on multivariate. Nevertheless, LSTM exhibited the most unfavorable

DOI: 10.55214/25768484.v8i5.1803

		SARIMA.			performance of all three strategies in terms of the sMAPE metrics.
[48]	LSTM and GRU	The forecasting of day-ahead EC in a hospital involved the use of LSTM and GRU networks with and without EMD and Complete Ensemble EMD preprocessing for both univariate and multivariate approaches.	24 hrs	24 hrs	In the comparison between LSTM and GRU models, both showed similar performance. However, incorporating EMD and Complete Ensemble EMD consistently enhanced results in the multivariate case. Notably, the best outcomes were achieved by LSTM with preprocessing in the multivariate scenario.
[49]	ERNN, seq2seq models, and temporal CNN	Aimed to bridge the gap by reviewing and experimentally evaluating four real-world datasets on the latest electric LF trends. Contrasting DL architectures (RNN, seq2seq models, and temporal CNN together with architectural variants) for STLF.	4 days	1 day	The Elman RNN performed comparably to GRU and LSTM when used for aggregated LF. Seq2seq models proved to be quite efficient in LF tasks, although they seem to fail to outperform RNNs. The temporal CNNs showed convincing performance in LF tasks. The LSTM performed best in multivariate aggregated LF. It is found that short-term LF at the customer level is also an extremely challenging task for DL models.

4. Data and Methodology

This section provides information on each dataset and discusses the forecasting stu-dies and results of these datasets as separate cases. The chapter concludes with a compa-rative discussion of the LSTM model's predictive results based on the datasets being con-sidered.

4.1. Common Methodology for All Cases

This study employed two distinct datasets with different structural characteristics for univariate and multivariate predictions. The forecasting process incorporated the use of the widely utilized LSTM model, known for its high forecasting performance potential. Prior to analysis, the datasets underwent data preprocessing, a technique for the transformation of raw data into a more relevant and clearer form. This involved data cleaning to remove missing values, and data transformation techniques to normalize the variables within a common interval, using a min-max scaler.

The data values range from 0 to 1. Outliers are values in a data set that significantly differ from the others. For electricity demand data, outliers could be due to public holidays like Christmas Day, or special events where energy demand is higher than usual. In this study, since we estimate the energy consumption for normal days during forecasting, the Z-score of such datasets is extracted using the Z-score technique, a technique widely used by researchers. After the data pre-processing phase is completed, the data is now ready to be transferred to the forecasting phase [15]. For the analysis of the data sets, the LSTM model employed the Keras DL library in the Python language [50]. The LSTM model is used to train on a specified sequence of EC, univariate input, at an interval that differs according to the datasets, and to forecast the output vector of the next step EC, which also differs according to the dataset. In multivariate input scenarios, the multivariate input is used to predict the next step's power consumption, which varies depending on the data set. The number of LSTM layers may vary for each case study in this study. To avoid overfitting, the number of hidden layers has been increased and followed by a dropout layer using a proper learning rate value for regularization [13]. The ReLU activation is used as it is the most effective with return sequences equal to true so that data can be passed from one layer to another [50]. The MSE is being utilized as the loss function for our specific DL model. We have opted for the Adam optimizer algorithm due to its status as an extension of stochastic gradient descent. Notably, the Adam optimizer has gained increased popularity in the realm of DL applications, particularly in fields such as CV and NLP. This wider adoption of the Adam optimizer is attributed to its effectiveness in training DL models. Finally, to assess the performance of the experimental study, we utilize the overall R-squared metric. R-squared represents the amount of variance in a dependent variable that's explained in a regression model by one or more independent variables. It provides insight into the goodness of fit of the model, with values ranging from 0 to 1. A score of 1 shows the model fits the data perfectly. This means that the closer this score is to 1, the better the prediction.

4.2. Case Studies and Forecasting Results

4.2.1. Data Case I

4.2.1.1. Data Set Description

The data set consists of aggregated data for Australia from January 1, 2006, to January 1, 2011, at a half-hourly interval. This dataset is resampled to daily intervals to avoid overtraining and consists of six time-dependent variables including electricity demand, electricity price, dew point temperature, humidity, dry bulb temperature, and wet bulb temperature. The EC values for the LSTM prediction model are used as the target variable.

4.2.1.2. LSTM Parameters

In the univariate case, a 64-unit LSTM layer is followed by 32-unit fully connected layers and an output layer. In the multivariate case, in contrast to the univariate case, a dropout layer that is regularized by a learning rate of 0.001 is followed by fully connected layers of 32 units. In both cases, the batch size is 57 and the training time is 40 epochs.

4.2.1.3. Learning Curves

The data set is divided into 80% for the training and 20% for the test. A good model fit is achieved when the model performs well on both training and validation sets without overfitting. The learning curve plots in **Figure 1**, Figure 2 depict the decreasing loss function over time and provide reliable forecasts for the model, which looks back 28 days for univariate and 14 days for multivariate. In this study, the blue-colored curves are the model training learning curves and the orange-colored curves are the model testing learning curves for each graph. When analyzing the training and testing datasets, these curves serve as a valuable tool for determining if the model is overfitted, underfitted, or fits appropriately. Upon observing the plots, it becomes evident that the loss function decreases swiftly to a low value, especially given the large training set. Losses decrease and stabilize around the same point during training and validation. Thus, the model successfully captures EC patterns.

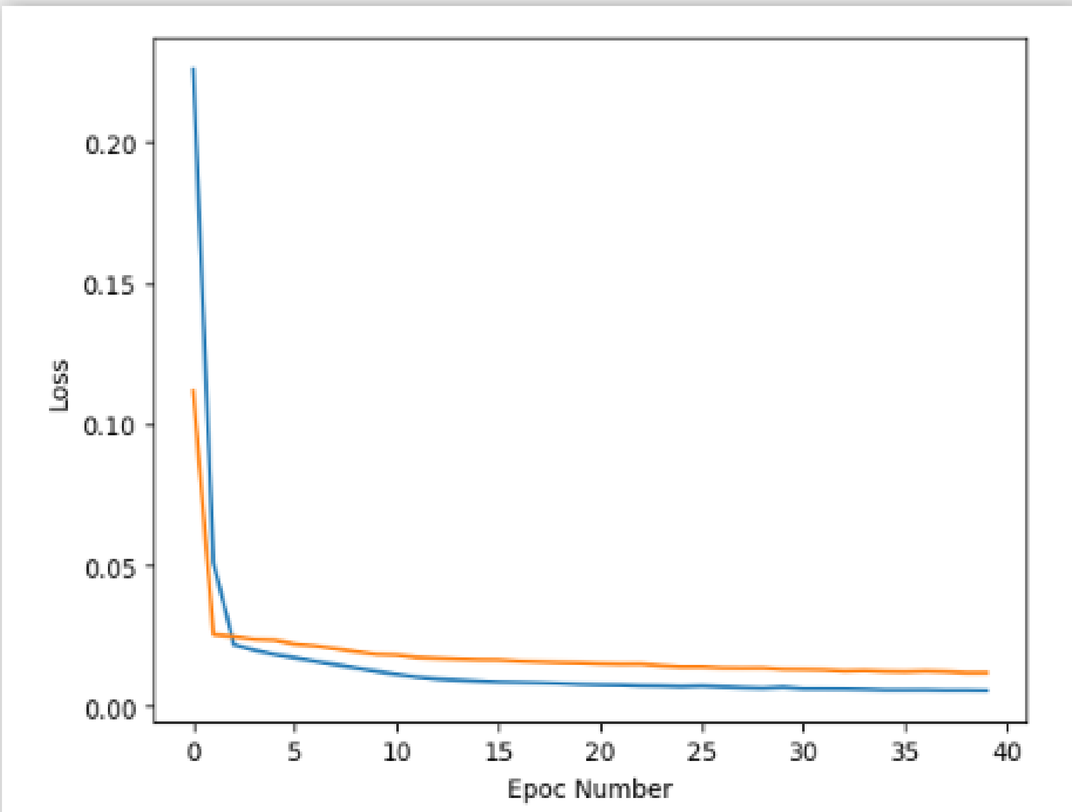

Figure 1.
Loss function decay of the training sets and testing sets for time series forecasting LSTM models of the univariate case by lagging 28 days.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

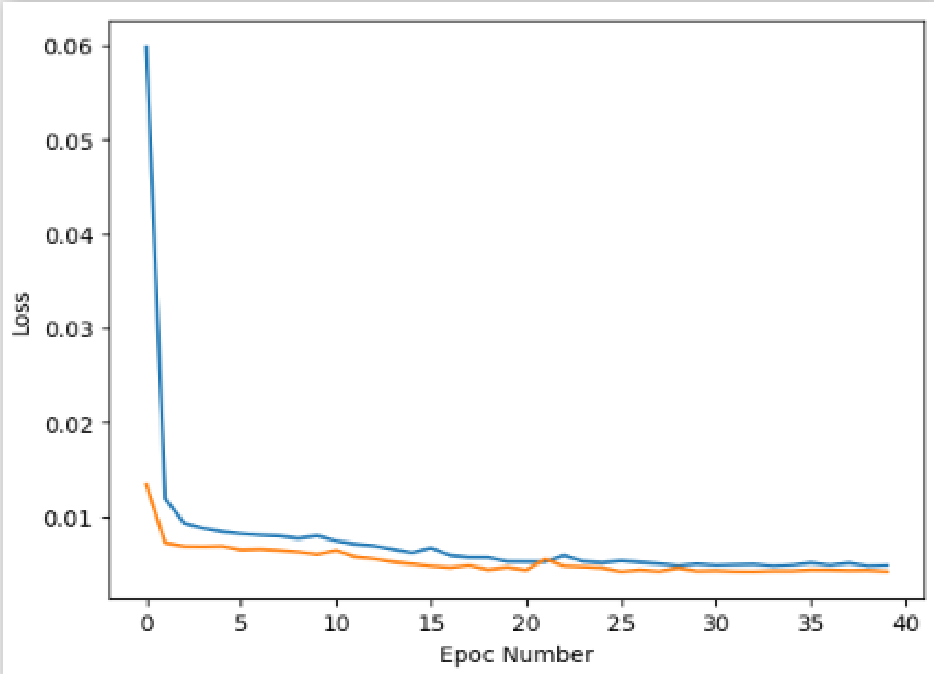

Figure 2.
Loss function decay of the training sets and testing sets for time series forecasting LSTM models of the multivariate case by lagging 14 days.

4.2.1.4. Forecasting results;

Table 4 below shows the estimation results obtained with this dataset. The values shown as steps in the table are historical lags, i.e. how many steps back (historical values) we look at when estimating the next value. For the daily time series data type, these values are often considered in the literature. In the LSTM forecasting model, the R-squared value is 0.7708 with the univariate time series dataset going back 7 days, while this value is 0.8209 with the multivariate time series data set, which is 6.5% higher than this value.

However, the R-squared values for the univariate and multivariate datasets are very close when looking at the 7, 14, 21, and 28 days back values. It is also noteworthy that while there is a significant increase in improving the R-squared scores for the predictive results obtained by looking back 7 and 14 days, there is no significant increase in predictive performance as the number of lagged days increases. In addition, we can determine which case has the best performance results by comparing the univariate and multivariate cases from 28 and 14 days ago.

Table 4.
LSTM model R-squared results for Data case I.

Lag period (day)	Univariate model	Multivariate model
7	0.7708	0.8209
14	0.8368	0.8534
21	0.8406	0.8487
28	0.8450	0.8519

Figure 3 and Figure 4 shows the actual power consumption curves and their prediction curves. It is evident from the figure that the accuracy increases in line with the test data.

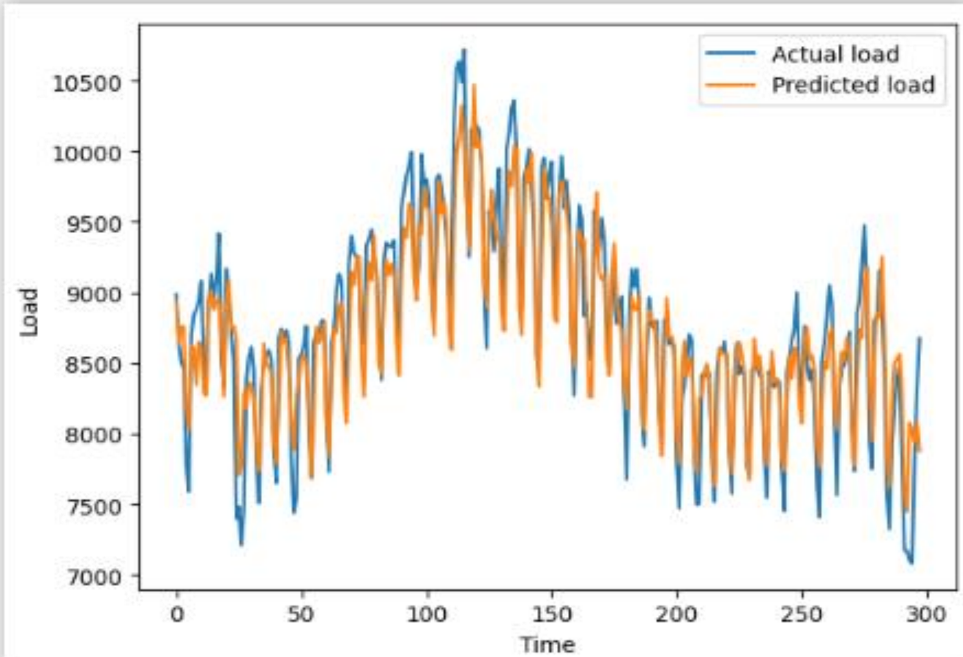

Figure 3.
The actual EC curves and its predictive curves of LSTM forecasting models by looking back periods of 28 days for univariate case.

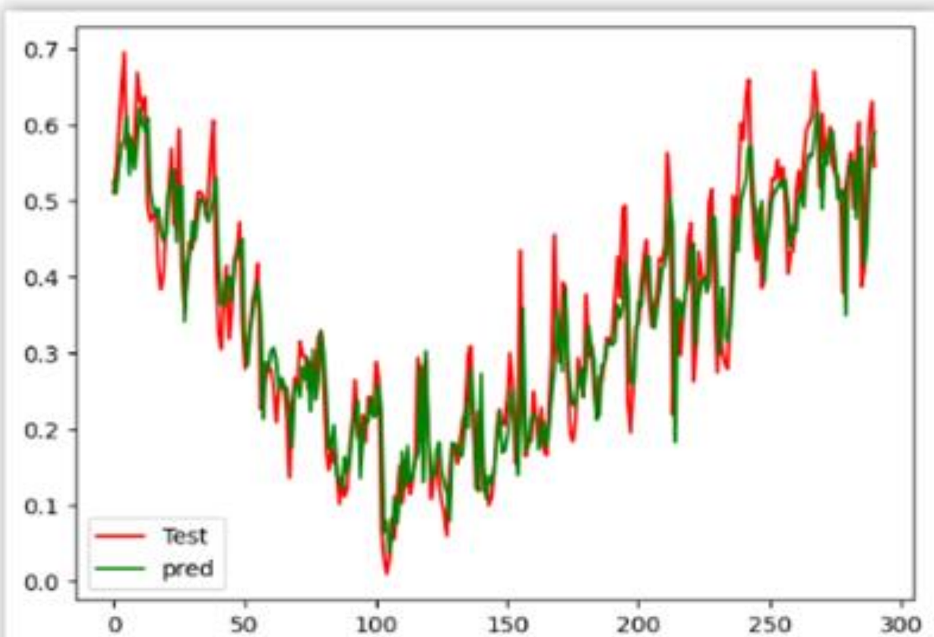

Figure 4.
The actual EC curves and its predictive curves of LSTM forecasting models by looking back periods of 14 days for multivarate case.

4.2.2. Data Case II

4.2.2.1. Dataset Description

For the typical evaluation procedure, a real dataset is used from the ISO New England control area and its eight wholesale load records for the year 2023. The wholesale records comprise Boston, Bridgeport, Burlington, Concord, Portland, Providence, Windsor Locks, and Worcester. This dataset has hourly intervals and three time-dependent variables: electricity demand, the dry bulb temperature, and the dew point temperature. The target variable is the EC values for the LSTM forecasting model.

4.2.2.2. LSTM Parameters

In the univariate case, an LSTM layer with 64 units is followed by a fully connected layer with 32 units each, a dropout layer regularized by 0.2, and an output layer. In the multivariate case, an LSTM layer with 64 units is followed by four fully connected layers with 32 units each followed by a separate layer of dropout regularized by 0.2, and an output layer. The batch size is set to 57 and the model is trained for 40 epochs for all cases.

4.2.2.3. Learning Curves

For training and testing, the data set is divided into 70% and 30% respectively. The graphs of the learning curves in Figure 5 and Figure 6 show the decrease of the loss function with epochs. These curves belong to the best predictions of the model obtained by looking back 48 hours for the univariate and 12 hours for the multivariate, respectively. The training and validation losses show a consistent decrease and stabilization at similar points, which indicates that the model has successfully captured the intricate patterns related to EC behaviors.

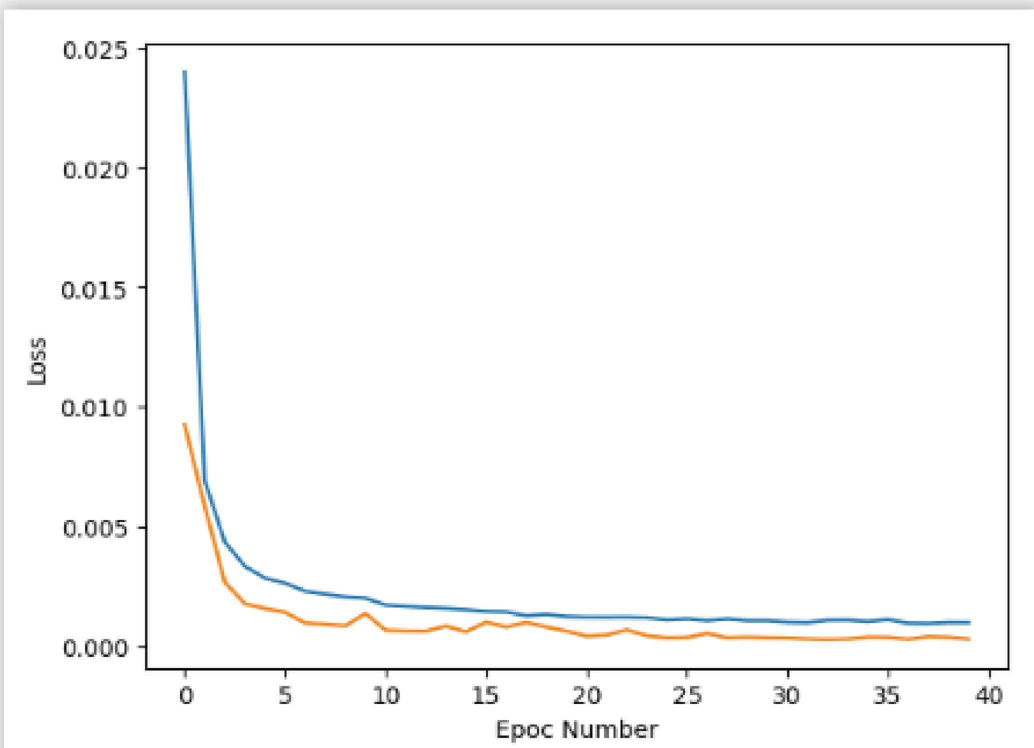

Figure 5.
Loss function decay of the training sets and testing sets for time series forecasting LSTM models of the univariate case by lagging 48 hours.

ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

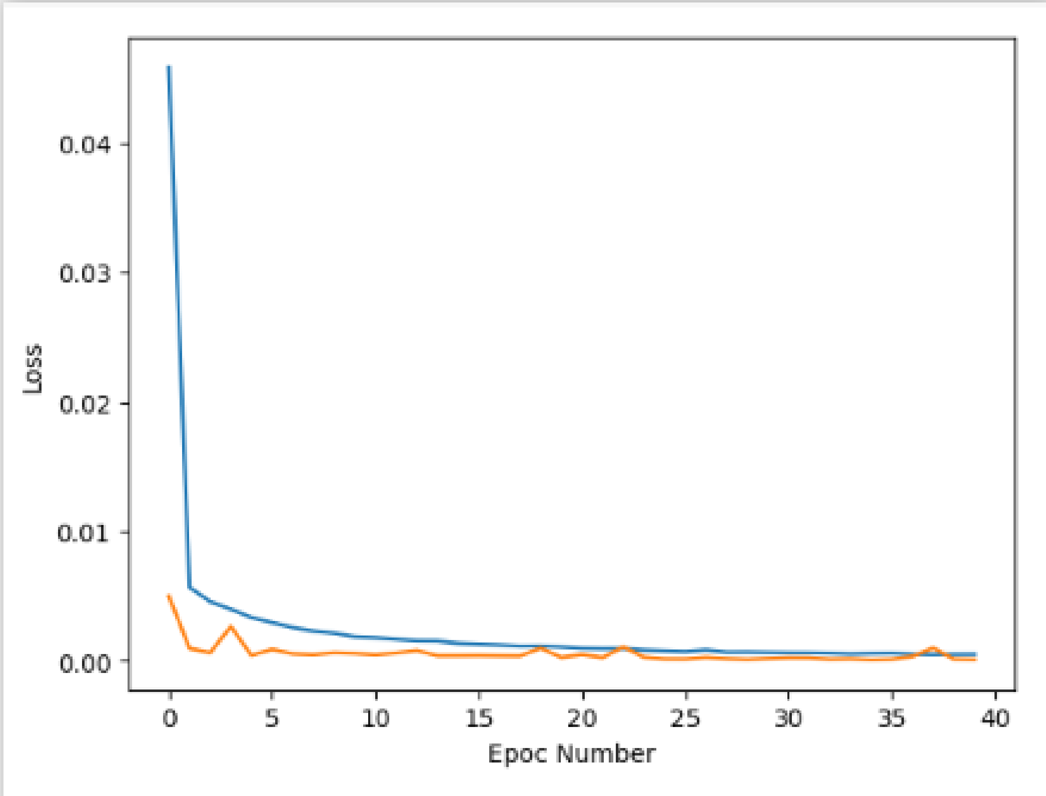

Figure 6.
Loss function decay of the training sets and testing sets for time series forecasting LSTM models of the multivariate case by lagging 12 hours.

4.2.2.4. Forecasting Results

With the hourly univariate and multivariate datasets, the EC value of the next hour is estimated by going back 6, 12, 24, and 48 hours. The estimation results obtained with the univariate dataset are high up to 48 hours while there is a small decrease in the R-squared score when looking back at 72 hours as seen in Table 5. However, R-squared scores of the estimation results obtained with the multivariate dataset show an increase up to 12 hours, while there is a significant decrease in this value when looking back at 24 and 48 hours. When Table 5 is analyzed in general terms, the R-squared scores of the estimation results obtained with univariate and multivariate datasets are quite high, but the R-squared scores of the estimation results obtained with univariate data sets are higher for each lag value examined. Furthermore, the actual EC curves and their predictive curves are shown in Figure 7 and Figure 8 by looking back at periods of 48 and 12 hours of univariate and multivariate cases which have the best performance results, respectively.

Table 5.
LSTM model R-squared results for Data case II.

Lag period(hr)	Univariate model	Multivariate model
6	0.9714	0.9530
12	0.9745	0.9642
24	0.9772	0.9365
48	0.9817	0.8761
72	0.9755	0.8640

ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

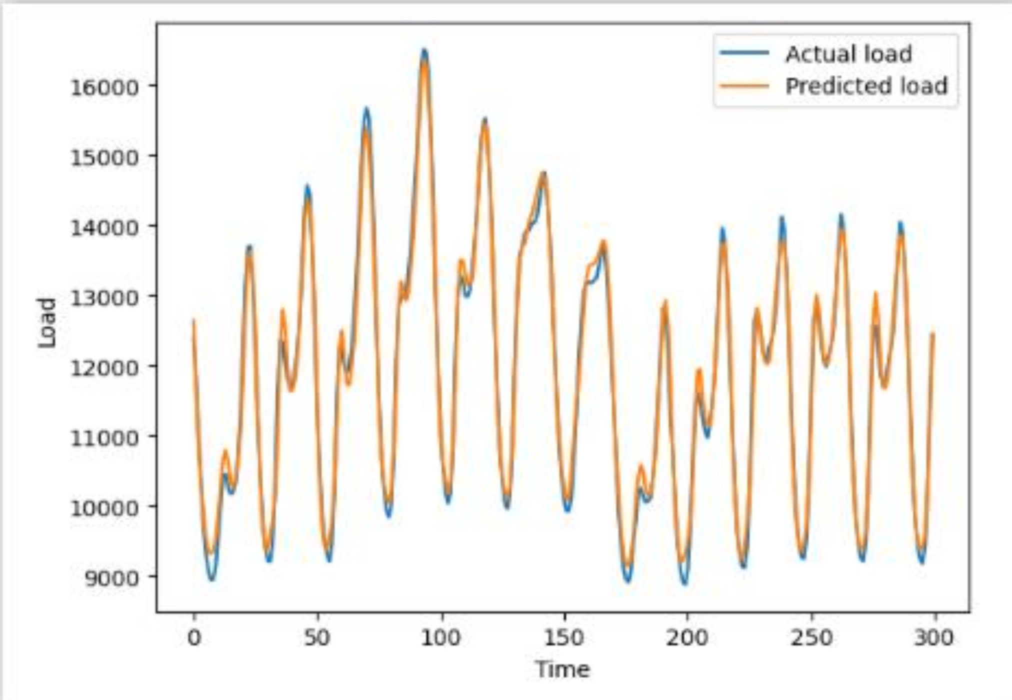

Figure 7.
The actual EC curves and its predictive curves of LSTM forecasting models by looking back at periods for 48 hours of univariate case.

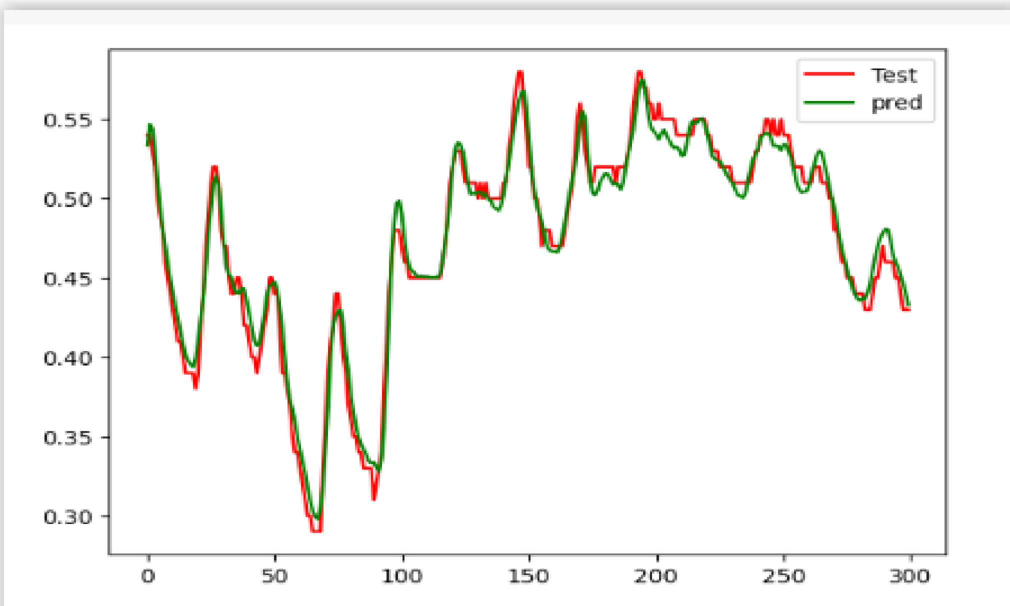

Figure 8.
The actual EC curves and its predictive curves of LSTM forecasting models by looking back at periods for 12 hours of multivarate case.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1014-1037, 2024
DOI: 10.55214/25768484.v8i5.1803

5. Discussion and Conclusion

According to the literature review, EC forecasts as multivariate time-series models are mostly short-term forecasts and are at the residential level, while EC forecasts as univariate time-series models are generally realized in short and medium-term periods and EC data is aggregated and at the system level. In both multivariate and univariate time series analysis studies, the simple LSTM model has shown much better EC forecasting performance compared to other ML-based models or statistical models, but its forecasting performance is lower than that of LSTM model variations. In addition, the hybrid or ensemble models, which are created by combining the LSTM model with other ML-based models, also perform better than the LSTM model variations. It is also noteworthy that hybrid and ensemble models are mostly constructed with combinations of LSTM-based approaches and CNN algorithms.

Based on our experimental evaluation, the most accurate forecasting results for each data case using LSTM models are summarized in Table 6 below. Upon reviewing the table, it is evident that in data case I, the univariate model achieves an R-squared score of 0.8450 with a 28-day lookback, while the multivariate model achieves a comparable score of 0.8534 with a 14-day lookback period. In data case II, we found R-square values very close to 1 when analyzing the previous 48 hours in the univariate model and 12 hours in the multivariate model. It's worth noting that the R-square score is higher in the univariate model compared to the multivariate model. The phenomenon occurs due to the way supervised learning is implemented in recurrent networks, such as deep LSTMs. In this approach, the neurons are randomly initialized, leading to the deactivation of neurons that are essential for accurately learning the latent features of the interrelated variables present in the multivariate time series dataset. This random initialization process can hinder the network's ability to effectively capture and understand the complex interdependencies within the dataset, thereby impacting its learning and predictive capabilities [13]. On the other hand, several papers have compared the results of multivariate and univariate time series analysis studies. They indicate that LSTM-based models are more effective in multivariate forecasting, as shown in Table 3. Nevertheless, it's worth noting that our experimental study, which utilized two distinct datasets, did not yield a conclusive determination on this matter. A closer assessment of our experimental study, presented in Table 6, reveals that the predictive performance of both univariate and multivariate forecasting models using hourly data significantly outperforms those using daily data. This suggests that as the time intervals of the datasets decrease, there is an observable improvement in LSTM forecasting results.

Table 6.
LSTM model R-squared results for data case II.

Cases	Steps		R-squared	
	Univariate	**Multivariate**	**Univariate**	**Multivariate**
Data case I	28 days	14 days	0,8450	0,8534
Data case II	48 hours	12 hours	0,9817	0,9642

Our upcoming research endeavors entail conducting an experimental study focused on the prediction of electricity prices, photovoltaic power output, and wind power output. Although there have been numerous multivariate and univariate time series analyses for electricity price forecasting in existing literature, there remains a noticeable gap in this specific context. Our aim is to bridge this gap by integrating these predictive studies with comprehensive power systems planning and management research to enhance the understanding of energy market dynamics and improve overall energy management strategies.

References

[1] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, R. Stoean, "On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Fmachine Models for Energy Load Forecasting," *Energies*, 16, 1434, 2023.
[2] Y. Wendong, S. Ismail, I. Hasnain, A. Sajid, "Modeling and Forecasting Electricity Demand and Prices: A Comparison of Alternative Approaches," *Journal of Mathematics*, 2022, 14, 2022.
[3] S. Bouktif, A. Fiaz, A. Ouni, M.A. Serhani, "Multi-Sequence LSTM-RNN Deep Learning and Metaheuristics for Electric Load Forecasting," *Energies*, 13, 391, 2020.
[4] M.A. Castán-Lascorz, P. Jiménez-Herrera, A. Troncoso, G. Asencio-Cortés, "A new hybrid method for predicting univariate and multivariate time series based on pattern forecasting," *Information Sciences* 586, 611-627, 2022. https://doi.org/10.1016/j.ins.2021.12.001
[5] C. M. Cheung, R. Kannan, V. K. Prasanna, "Temporal ensemble learning of univariate methods for short term load forecasting, "*IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA, 19-22 Feb. 2018.
[6] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, A. Troncoso, "Deep Learning for Time Series Forecasting:A Survey," Big Data, 9, 3-21, 2021.
[7] A. Casolaro, V. Capone, G. Iannuzzo, F. Camastra, "Deep Learning for Time Series Forecasting: Advances and Open Problems," Information, 14, 598, 2023.
[8] Aseeri, A. O. , "Effective RNN-Based Forecasting Methodology Design for Improving Short-Term Power Load Forecasts: Application to Large-Scale Power-Grid Time Series," Journal of Computational Science, 68, 101984, 2023.
[9] Petropoulos, F.; Spiliotis, E. , "The Wisdom of the Data: Getting the Most Out of Univariate Time Series Forecasting," Forecasting, 3, 478-497, 2021.
[10] Bilgili, M.; Engin, P. , "Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye," Energy, 284(C), 128575, 2023.
[11] Zhao, Q.; Yang, G.; Zhao, K.; Yin, J.; Rao W.; Chen, L. , "Multivariate Time-Series Forecasting Model: Predictability Analysis and Empirical Study," IEEE Transactions on Big Data, 9,1536-1548, 2023.
[12] Yang, Y.; Lu, J. , "Foreformer: an enhanced transformer-based framework for multivariate time series forecasting," Appl Intell, 53, 12521–12540, 2023.
[13] Sagheer, A.; Kotb, M. , "Unsupervised Pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate Time Series Forecasting Problems,"Sci Rep., 9, 19038, 2019.
[14] Nespoli, A.; Ogliari, E.; Pretto, S.; Gavazzeni, M.; Vigani, S.; Paccanelli, F. , "Electrical Load Forecast by Means of LSTM: The Impact of Data Quality," Forecasting, 3, 91-101, 2021.
[15] Khalid, R.; Javaid, N.; Al-zahrani, F.A.; Aurangzeb, K.; Qazi, E.-u.-H.; Ashfaq, T. , "Electricity Load and Price Forecasting Using Jaya-Long Short Term Memory (JLSTM) in Smart Grids," Entropy, 22, 10, 2020.
[16] Alonso, A.M.; Nogales, F.J.; Ruiz, C. , "A Single Scalable LSTM Model for Short-Term Forecasting of Massive Electricity Time Series," Energies, 13, 5328, 2020.
[17] Dubey, A. K.; Kumar, A.; García-Díaz, V.; Sharma, A. K.; Kanhaiya, K. , "Study and analysis of SARIMA and LSTM in forecasting time series data," Sustainable Energy Technologies and Assessments, 47, 101474, 2021.
[18] Kim, T.; Cho, S. , "Predicting residential energy consumption using CNN-LSTM neural networks," Energy, 182, 72-81, 2019.
[19] Chung, J.; Jang, B. , "Accurate prediction of electricity consumption using a hybrid CNN-LSTM model based on multivariable data," PLoS One, 17(11), 0278071, 2022.
[20] Mujeeb, S.; Javaid, N.; Ilahi, M.; Wadud, Z.; Ishmanov, F.; Afzal, M.K. , "Deep Long Short-Term Memory: A New Price and Load Forecasting Scheme for Big Data in Smart Cities," Sustainability, 11, 987, 2019.
[21] Son, H.; Kim, C. , "A Deep Learning Approach to Forecasting Monthly Demand for Residential–Sector Electricity," Sustainability 12, 3103, 2020.
[22] Li, K.; Huang, W.; Hu, G.; Li, J. , "Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network," Energy and Buildings, 279, 112666, 2023.
[23] Du, S.; Li, T.; Yang, Y.; Horng, S. , "Multivariate time series forecasting via attention-based encoder–decoder framework," Neurocomputing, 388, 269-279, 2020.
[24] Bhoj, N.; Bhadoria, R. S. , "Time-series based prediction for energy consumption of smart home data using hybrid convolution-recurrent neural network," Telematics and Informatics, 75, 101907, 2022.
[25] La Tona, G.; Luna, M.; Di Piazza, M.C. , "Day-ahead forecasting of residential electric power consumption for energy management using Long Short-Term Memory encoder–decoder model," Mathematics and Computers in Simulation, 2023.
[26] Das, A.; Annaqeeb, M. K.; Azar, E.; Novakovic, V.; Kjærgaard, M. B. , "Occupant-centric miscellaneous electric loads prediction in buildings using state-of-the-art deep learning methods," Applied Energy, 269, 115135, 2020.
[27] Rafi, S. H.; Masood, N. A.; Deeba, S. R.; Hossain, E. , "A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network," IEEE Access, 9, 32436-32448, 2021.

[28] Farsi, B.; Amayri, M.; Bouguila, N.; Eicker, U. , "On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach," IEEE Access, 9, 31191-31212, 2021.
[29] Hou, T.; Fang, R.; Tang, J.; Ge, G.; Yang, D.; Liu, J.; Zhang, W. , "A Novel Short-Term Residential Electric Load Forecasting Method Based on Adaptive Load Aggregation and Deep Learning Algorithms," Energies, 14, 7820, 2021.
[30] Ozer, I.; Efe, S. B.; Ozbay, H. , "A combined deep learning application for short term load forecasting," Alexandria Engineering Journal, 60, 3807-3818, 2021.
[31] Sagheer, A.; Kotb, M. , "Unsupervised Pre-training of a Deep LSTM-based Stacked Autoencoder for Multivariate Time Series Forecasting Problems," Sci Rep., 9, 19038, 2019.
[32] Shaqour, A.; Ono, T.; Hagishima, A.; Farzaneh, H. , "Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building," Energy and AI, 8, 100141, 2022.
[33] Le, T.; Vo, M.T.; Vo, B.; Hwang, E.; Rho, S.; Baik, S.W. , "Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM," Appl. Sci., 9, 4237, 2019.
[34] Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. , "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †," Energies, 11, 1636, 2018.
[35] Hadjout, D.; Torres, J.F.; Troncoso, A.; Sebaa, A.; Martínez-Álvarez, F. , "Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market," Energy, 243, 123060, 2022.
[36] Pierre, A.A.; Akim, S.A.; Semenyo, A.K.; Babiga, B. , "Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA-LSTM and ARIMA-GRU Approaches," Energies, 16, 4739, 2023.
[37] Masood, Z.; Gantassi, R.; Ardiansyah; Choi, Y. , "A Multi-Step Time-Series Clustering-Based Seq2Seq LSTM Learning for a Single Household Electricity Load Forecasting," Energies, 15, 2623, 2022.
[38] Silva, D. G.; Moura Meneses, A. A. , "Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction," Energy Reports, 10, 3315-3334, 2023.
[39] Torres, J.F.; Martínez-Álvarez, F.; Troncoso, A. , "A deep LSTM network for the Spanish electricity consumption forecasting," Neural Comput & Applic, 34, 10533–10545, 2022.
[40] Ünlü, K.D. , "A Data-Driven Model to Forecast Multi-Step Ahead Time Series of Turkish Daily Electricity Load," Electronics, 11, 1524, 2022.
[41] Kong, W.; Dong, Z. Y.; Jia, Y.; Hill, D. J.; Xu Y.; Zhang, Y. , "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," IEEE Transactions on Smart Grid, 10, 841-851, 2019.
[42] Yan, K.; Li, W.; Ji, Z.; Qi, M.; Du, Y. , "A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households. IEEE Access, 7, 157633-157642, 2019. https://doi.org/10.1109/ACCESS.2019.2949065.
[43] Chaturvedi, S.; Rajasekar, E.; Natarajan, S.; McCullen, N. , "A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India," Energy Policy, 168, 113097, 2022. Gul, M.J.; Urfa, G.M.; Paul, A. et al. , "Mid-term electricity load prediction using CNN and Bi-LSTM," J Supercomput, 77, 10942–10958, 2021.
[44] Gul, M.J.; Urfa, G.M.; Paul, A. et al. Mid-term electricity load prediction using CNN and Bi-LSTM. J Supercomput 2021, 77, 10942–10958.
[45] Ghimire, S.; Deo, R. C.; Casillas-Pérez, D., Salcedo-Sanz, S. , "Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach," Energy Conversion and Management, 297, 117707, 2023.
[46] Son, N. , "Comparison of the Deep Learning Performance for Short-Term Power Load Forecasting," Sustainability, 13, 12493, 2021.
[47] Hadri, S.; Najib, M.; Bakhouya, M.; Fakhri, Y.; El Arroussi, M. , "Performance Evaluation of Forecasting Strategies for Electricity Consumption in Buildings," Energies, 14, 5831, 2021.
[48] Fernández-Martínez, D.; Jaramillo-Morán, M.A. , "Multi-Step Hourly Power Consumption Forecasting in a Healthcare Building with Recurrent Neural Networks and Empirical Mode Decomposition," Sensors, 22, 3664, 2022.
[49] Gasparin, A.; Luković, S.; Alippi, C. , " Deep learning for time series forecasting: the electric load case," CAAI Transactions on Intelligence Technology, 7(1), 1-25, 2021.
[50] Shin, S.-Y.; Woo, H.-G. , "Energy Consumption Forecasting in Korea Using Machine Learning Algorithms," Energies 15, 4880, 2022.

Appendix 1.

Nomenclature			
LSTM	Long-short term memory	MLR	Multiple linear regression
Bi-LSTM	Bi-directional LSTM	SVR	Support vector regression
DLSTM	Deep LSTM	FFNN	Feed-forward neural network
DL	Deep learning	D-FFNN	Deep-FFNN
EC	Electricity consumption	CEEMDAN-SE-LSTM	Complete ensemble empirical mode decomposition with adaptive noise sample Entropy LSTM

ML	Machine learning	PVGIS	Photovoltaic geographical information system
ReLU	Rectified linear units	RBFN	Radial basis functional network
CNN	Convolutional neural network	XGBoost	Extreme gradient boosting
AR	Autoregressive	BPNN	Back-propagation neural network
ARMA	AR moving average	OPTICS	Ordering points to identify the clustering structure
ARIMA	AR integrated moving average	XCORR	Cross-correlation
SARIMA	Seasonal ARIMA	LGBM	Light gradient boosting machine
ARIMAX	ARIMA with explanatory variables	GA	Genetic algorithm
NARX	Nonlinear AR with exogenous inputs	PSO	Particle swarm optimization
ARCH	AR conditional heteroskedasticity	RF	Random forest
GARCH	Generalized ARCH	DT	Decision tree
SGARCH	Symmetric GARCH	GDP	Gross domestic product
TGARCH	The threshold GARCH	FCL	Fully connected layers
EGARCH	Exponential GARCH	DNN	Deep neural network
ANN	Artificial neural network	TCN	Temporal convolution networks
LF	Load Forecasting	Seq2Seq	Sequence-to-sequence
RNN	Recurrent neural network	GRN	Gated recurrent network
ReLU	Rectified linear units	MSA	Metaheuristic search algorithms
UV	Ultra viole	SWT	Stationary wavelet transform
GRU	Gated recurrent unit	ICMD	Improved complete ensemble empirical mode decomposition with adaptive Noise
CNN	Convolutional neural network	EMD	Empirical mode decomposition
ELM	Extreme learning machine	CV	Computer vision
MAE	Mean absolute error	NLP	Natural language processing
MAPE	Mean absolute percentage error		
sMAPE	Symmetric MAPE		
RMSE	Root mean squared error		
NRMSE	Normalized RMSE		
R-squared	Coefficient of determination		