



Article

Machine Learning Short-Term Energy Consumption Forecasting for Microgrids in a Manufacturing Plant

Maciej Slowik *  and Wieslaw Urban 

Faculty of Engineering Management, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Poland; w.urban@pb.edu.pl

* Correspondence: m.slowik@pb.edu.pl

Abstract: Energy production and supply are important challenges for civilisation. Renewable energy sources present an increased share of the energy supply. Under these circumstances, small-scale grids operating in small areas as fully functioning energy systems are becoming an interesting solution. One crucial element to the success of micro-grid structures is the accurate forecasting of energy consumption by large customers, such as factories. This study aimed to develop a universal forecasting tool for energy consumption by end-use consumers. The tool estimates energy use based on real energy-consumption data obtained from a factory or a production machine. This model allows the end-users to be equipped with an energy demand prediction, enabling them to participate more effectively in the smart grid energy market. A single, long short-term memory (LSTM)-layer-based artificial neural network model for short-term energy demand prediction was developed. The model was based on a manufacturing plant's energy consumption data. The model is characterised by high prediction capability, and it predicted energy consumption, with a mean absolute error value of 0.0464. The developed model was compared with two other methodologies.

Keywords: short-term forecasting; energy consumption; microgrids; smart grids; LSTM



Citation: Slowik, M.; Urban, W. Machine Learning Short-Term Energy Consumption Forecasting for Microgrids in a Manufacturing Plant. *Energies* **2022**, *15*, 3382. <https://doi.org/10.3390/en15093382>

Academic Editor: Tiago Pinto

Received: 13 April 2022

Accepted: 3 May 2022

Published: 6 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Energy production and supply are undoubtedly crucial issues in modern economies. Most countries are transforming their energy systems towards cleaner, greener, and entirely renewable alternatives. Meeting civilisation's demand requires different approaches to link energy systems, production, transmission, and consumption. This challenge is typical when shifting from highly centralised energy production systems toward diffuse production systems, small-scale energy production systems, and microgrids. Furthermore, renewable energy sources are characterised by changeable production capacity, where variables are difficult to predict accurately. In such circumstances, forecasting, including short-term forecasting, in various nodes of an energy system becomes a vital and scientifically critical issue.

A microgrid is a small-scale grid. However, it is fully functional, operating in a limited geographical area; it can operate independently or be connected to a larger grid [1,2]. Worighi et al. [3] presented a smart-grid architecture model. The model is a smart grid consisting of the main grid and multiple embedded microgrids. The authors underline the benefits of small-scale grids that can maximise the local resources and reduce economic and energy losses during the power transmission. They can supply power stability, shift peak-load demand, and lower carbon emissions [2] (p. 238). The benefits are multidimensional; consumers can control their energy usage in a more flexible, reliable manner while considering economic factors at the same time [4].

Microgrids complement demand response programs. The demand response relies upon consumers actively involved in a smart grid, and they can adjust their electricity consumption during peak times and may benefit financially [5]. Such a program implies

increasing and reducing energy demand in everyday use; it also refers to load curtailment through signals sent by the supplier or grid operator [1]. Alotaibi et al. [5] indicated that *the participation of demands* is based on economic incentives for consumers arising from the dynamic pricing framework. Several dynamic pricing modalities have been implemented for demand response, such as time-of-use, critical peak pricing, real-time pricing, and day-ahead pricing [5] (p. 19). Various small-scale components, including demand response, challenge the microgrid's modelling and planning. As Shulyma et al. [6] stated, the microgrid system is characterised by rapidly changing operating modes and configurations.

Smart grids and their applications vary in form and function. Palma-Behnke et al. [7] introduced a design for a smart microgrid solution for rural, isolated towns in Chile. Masembe [8] proposed smart grid technologies as a solution for increasing reliability and reducing power outages. Samad et al. [9] outlined varied smart grid technologies with industrial use case studies. Keller et al. [10] presented future interaction methods of factories and smart grids and manufacturing integration. The common approach described above is that smart grids have modularity, flexibility, and the ability to use different power sources—renewable and fossil fuels. Smart grids can be stand-alone—as so-called islands or connect with energy grids. To operate, elements of smart grids must be capable of communicating. One such method is machine to machine (M2M) to exchange measurement, failure, and diagnostic information. Another method is high-level communication, performed by humans or artificial agents, connected with decision making about buying, selling, and bidding energy on automated markets.

The novelty and the main contribution of this study is a developed method of forecasting energy consumption by the end-user (factory or machine), along with an example application. The proposed prediction model was based on long short-term memory (LSTM) network which allows for using measured energy consumption so far. The proposed solution makes it possible to predict energy requirements and, based on that, take part as a consumer or prosumer in the local smart grid or other energy markets. In contrast to most researchers connected with forecasting energy demand for the prognosis of the entire national energy market, our proposed solution focused on the end-user perspective, and forecasting was based on real-sensor-based measurements.

2. Forecasting Methods for Smart Grids

Most researchers focus on the microgrid or smart grid as a larger system—or a scalable system of systems, e.g., Shahid [11]—in which renewable energy sources are used as power generators. We can classify groups based on photovoltaic (PV) forecasts and solar irradiance (W/m^2) to estimate cloud structure over time. Proper scheduling and planning of power system operations need to have an accurate load demand and renewable energy generation estimation from many sources, particularly in the short term, e.g., an hour ahead or a day ahead [12]. Therefore, the issue of forecasting concerning smart grids is presented by the literature as a crucial outcome [2,4,13,14]. Forecasting helps manage the produced energy in a dispatchable manner [4]. According to Ma and Ma [2] (p. 241), the major forecasting techniques used in that field can be categorised as statistical parametric or nonparametric intelligent methods and a hybrid model.

In recent years, scholars have assessed varied strategies and algorithms of forecasting for smart grids. Ahmad et al. [15] proposed a hybrid artificial neural network-based day-ahead load-forecasting model for smart electricity grids. The authors underlined the importance of short-term customer load forecasting in smart grids; it might impact decisions, such as generating capacity scheduling and energy transaction planning [15]. Yaprakdal et al. [12] elaborated bi-directional LSTM units based on a deep recurrent neural network model to reduce power losses by optimal operational scheduling of reconfigurable microgrids. Wood [14] applied machine-learning methods for power time-series predictions in the case of the German electricity market. Wang et al. [16] proposed operation optimisation modelling for microgrids considering distributed generation, environmental factors, and demand response. Kempener et al. [17], in a guiding document, also under-

lined the concern of one-to-six-hour-ahead predictions for smart grids. They called them *nowcasting* because of the short periods considered.

Studies in the literature classify forecasting techniques in many ways. Two general approaches need to be considered for this study design. The first type of forecasting technique is classical methods. They are based on process modelling and statistical analysis of parameters such as autocorrelation, seasonality, and trends. With classical techniques based on mathematical process modelling and analysis of time-series data, we can differentiate between other groups, e.g., autoregressive integrated moving average (ARIMA) [18], autoregressive moving-average models with exogenous input (ARMAX) [19] or seasonal ARIMA (SARIMA) [20]. Another group of forecasting models is based on identification and mathematical modelling, e.g., state-space models. The different Kalman filter modifications can be separated into extended or unscented Kalman filters. Using processing and measurement equations and external sensor data with chosen weights, we can estimate the future state of a modelled object. Su et al. [21] and Emami et al. [22] used modified Kalman filters for short-term wind speed and traffic flow predictions. Another approach [23] considers parametric and nonparametric model usage. The authors of [24] presented also sequential models for the time-evolving process of energy demand prediction. Lisi and Shah [25] showed several functional models of next-day forecasting of energy consumption. Basis functions for energy prediction in the Japan Electric Power Exchange and LASSO method for choosing optimal parameters were shown by Hirose et al. [26]. Ensemble-based model for forecasting spot prices in the Italian electricity market was described by Bibi et al. [27]. In addition, another study presented ARMA-based forecasting of the Pakistan electricity consumption data [28].

In-depth studies identified the two following approaches to forecasting, with a few crucial limitations:

- (a) Requirements and constraints connected with mathematical modelling. Difficulties related to the requirements for accurate mathematical description due to linearity requirements and data approximation;
- (b) Knowledge of parameters and characteristics, e.g., a lack of sensor parameter limitations, including measurement accuracy from which data are collected (to be used in determining the confidence weights for the measurement equation of the Kalman filter);
- (c) Requirements to fulfil linearity, stationarity, and seasonality assumptions, requiring earlier time series analysis. Knowledge of what is feasible or connected with various sources of data uncertainties;
- (d) Restrictions from processes and errors connected with measured values in real-life scenarios, e.g., errors in data series, empty values, time series parameters such as nonlinearity, seasonality, or trend.

Beyond these classical methods, a group of methods is based on machine learning techniques. We can distinguish techniques such as support vector machines (SVMs), convolutional neural networks (CNNs), and LSTM. The specificity of these methods was further analysed, considering their use related to the subject of these studies by others. Zhang et al. [29] used the SVM technique for PV fault detection problem prediction. Artificial neural networks of diverse types and architectures are often used in topics related to energy forecasting; developing a neural network model trains the network to recognise patterns based on input and output data. The first category of the artificial neural network is CNN, the main application of which is image recognition. Del Real et al. [30] described a CNN model that recognised energy demand on the French power grid. The second category is a recurrent neural network (RNN) model, commonly used in speech and text recognition. Kang et al. [31] compared CNN and RNN and hybrid models in forecasting power demands using the Korean energy grid data. Kumar et al. [32] chose models based on LSTM and gated recurrent unit (GRU) networks to forecast household electricity consumption. Mele et al. [33] illustrated self-organising maps and k-nearest neighbours (KNNs) as other examples of machine learning techniques adopted in energy forecasting.

Konstantinou et al. [34] proposed a prediction model based on the LSTM artificial neural network to predict power levels generated by PV stations over 1.5 h ahead. Ahn et al. [35] proposed a PV forecast RNN-based short-term algorithm based on measurements of power internet of things (IoT) sensors. Brahma et al. [36] proposed deep neural networks to create a daily forecasting model to predict solar irradiance based on a historical dataset with a range of 36 years. Jeon et al. [37] suggested a different approach, using the Korean meteorological service's next-day weather forecasts (i.e., temperature, humidity, and wind speed). Four different methods in the scope of forecasting electricity energy consumption were compared by Bilgili et al. [38]. Pen et al. [39] used wavelet transform to increase the accuracy of LSTM networks. Adapting a genetic algorithm to choose optimal hyperparameters for forecasting the LSTM network was also presented [40].

Pena-Gallardo et al. [41] showed that machine learning methods have better performance than ARIMA but need more time for training and retraining, whereas ARIMA cofactors are computed only once. Pao [42] compared an artificial neural network (ANN) with the ARMAX model to predict energy consumption. The analysis shows that the ANN model is better at identifying nonlinearities and nonlinear effects. However, despite the growing popularity of smart grids and connected technologies, the literature shows that existing models and techniques are case-specific and focus on forecasts of energies generated by smart grids or renewable energy sources because different factors relate to creating prediction models based on neural networks. We identify problems with obtaining datasets, features engineering, measurement quality, computation time, and solutions quality.

Different research groups have used LSTM-based models and their different variants to forecast energy consumption. Wang et al. [43] studied the LSTM network in terms of its application for periodic time-series prediction. Their model's prediction performance was higher than traditional forecasting methods based on a statistical approach, e.g., autoregressive moving average (ARMA) or autoregressive fractional integrated moving average (ARFIMA) models. The paper also emphasised LSTM's capacity for generalisation. Laib et al. [44] proposed a hybrid architecture with an LSTM network for the prediction of natural gas consumption, with a multilayer perceptron neural network.

Among the disadvantages, Deligiannidis et al. [45] showed that RNNs could have comparable prediction performance but with a smaller number of units. Additionally, Kaur et al. [46] showed that LSTM-based solutions could be inferior to RNN-based networks in smaller datasets. Choi et al. [47] proposed a different approach for forecasting, by using ANN models with an ensemble of different LSTM networks, resulting in adaptive weighting. Ahmed et al. [48] showed the limitations of LSTM networks in comparison to techniques for PV solar power forecasting. Then, they proposed their derivatives, named deep LSTM, and different solutions of deep convolutional neural networks (DCNNs), as future alternatives.

Sun et al. [49] compared different forecasting methods in building energy prediction. Among others, ANN has a higher percentage of use among considered studies. However, in second place are support vector regression (SVR) models. The third place is occupied by models based on ensemble methods. In the fourth position are deep learning models. Somu et al. [50] proposed kCNN-LSTM, a deep learning framework for building energy demand predictions. Luo et al. [51] presented a different approach—a physically constrained LSTM to limit models from generating unrealistic results.

This study aimed to develop a universal forecasting tool intended for energy consumption forecasts by end-use consumers. The developed tool estimates energy use based on real data and energy used by specific equipment. The forecasting model provides end-users with information on energy use prognosis, allowing them to participate in the smart grid energy market in an economically beneficial manner.

The developed solution allows the end-users to forecast factory energy requirements from minutes to four hours ahead, using the previous 24 h supply data. It combines solutions to two opposing problems using neural networks, multiple datasets for the learning process, and the complicated design of ANN architecture. Expensive, proprietary

software packages dedicated to scientific research were not used for the program, making it more feasible to implement. The model can use only one variable to forecast data prediction, which simplifies its use and makes it practical in industrial environments. The systematic summary of methods employed by scholars in the field of energy forecasting is presented in Table 1 below.

Table 1. The summary of research studies associated with energy forecasting.

References	Main Area of Work Described in Paper	Methods/Algorithms Used
[2,18,19]	Summary of forecasting methods for microgrids	ARIMA, ARMAX, ANN, others
[4,29]	Weather prediction for microgrids	Artificial intelligence methods
[12,31–33]	Short-term operation management and forecasting	Deep neural networks, ANN
[13]	Forecasting for Smart Cities	Hybrid algorithms
[14,34–37]	Wind and solar energy prediction	Transparent open box algorithm
[18,38,41]	ARIMA vs. LSTM	LSTM
[21,22]	Modelling of electricity consumption in changing environment	Kalman filters, Model Predictive Control
[23–28]	Electricity market forecasting	Parametric and nonparametric approach, statistical methods

3. Proposed Methodology

3.1. Proposed Machine Learning Model and Data Characteristics

The proposed machine learning model needs to take into consideration the following limitations:

- (1) Energy use data—the time range for which forecasting is made;
- (2) Application for the end-user—the model needs to forecast energy consumption without the employment of expensive software, and therefore, the proposed solution is a utility that is easy to use and does not require knowledge in forecasting techniques;
- (3) The character of the measurement data—the shape of the input vector and the number of data points impact the preparation of training and test datasets;
- (4) Automation of energy prediction forecasting.

3.1.1. Dataset Exploited in this Study

The dataset used for the energy consumption forecast was acquired by recording energy use in a manufacturing plant. The measurements were recorded with PQ BOX 100, a portable analyser of power quality parameters. The energy (W) measurements used in the proposed model were 8640 samples taken at intervals of 10 s over 24 h (360 per hour). The basic statistical characteristics of the data are described in Table 2, and the values are shown in Figure 1. The presented measurements show the total electricity used by all pieces of equipment in the manufacturing plant, which is located in the northeast of Poland.

Table 2. Basic characteristics of the dataset.

Number of samples	8640
Mean value	122,241.410
Standard deviation	64,964.959
Minimum value	20,386.545
Maximum value	241,949.547
Q1	72,733.056
Median	82,668.418
Q3	203,871.187

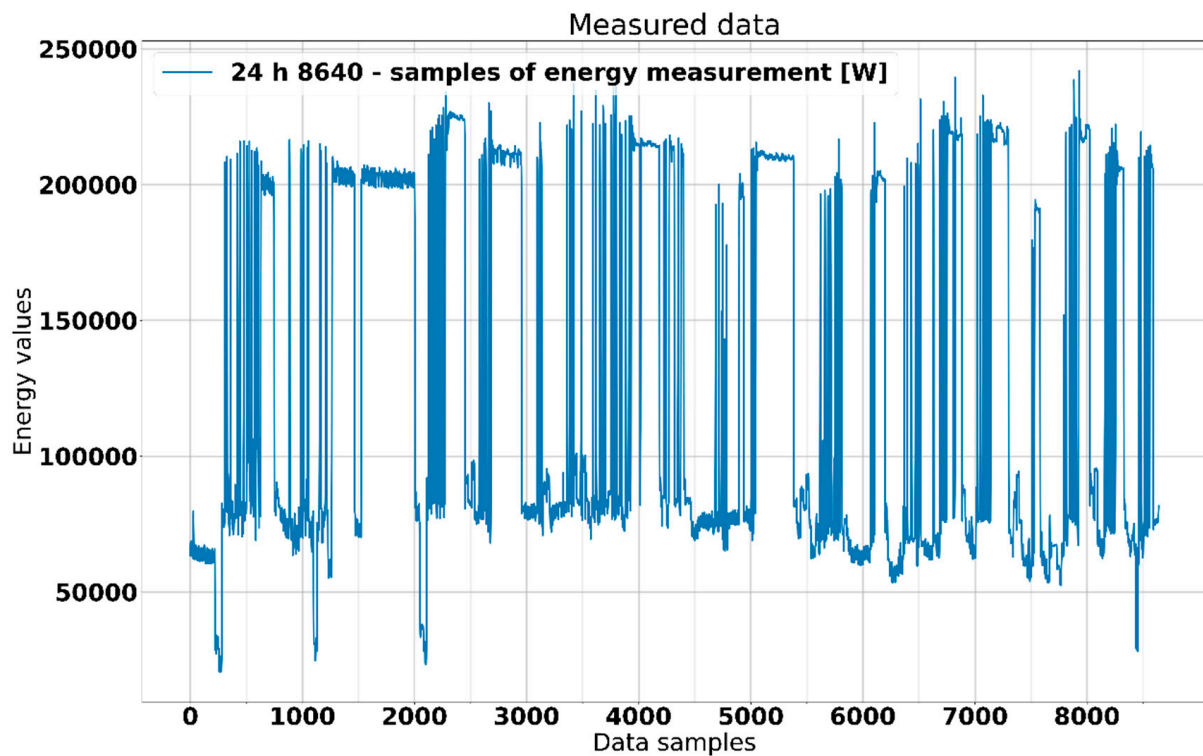


Figure 1. Measurement data graph—24 h.

Figure 1 presents highly nonlinear energy measurements with rapid changes in values and a broad range of amplitude.

3.1.2. Machine Learning Model

Due to the limitations and requirements described in Section 2 and the character of measurement data, the authors considered using an RNN—a class of ANN that has an internal memory state and the ability to process past time inputs. The disadvantage of RNN is that this group of ANN can only *remember* short-term information. Manowska [52] presented LSTM neural networks as a type of RNN with a feedback loop. This feature allows them to overcome temporary memory barriers and use them as a default choice in forecasting models. LSTM, described as such by Hochreiter et al. [53] due to the above properties, was also used in handwriting recognition tasks [54]. Sak et al. [55] used LSTM as architecture for large-scale acoustic modelling. Wu et al. [56] defined Google’s Neural Machine Translation (GNMT) system as composed of a deep LSTM network with eight encoder and decoder layers. The LSTM unit can be defined by Equations (1)–(6) as follows:

$$f_t = \sigma_g(W_f \times x_t + U_f \times h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma_g(W_i \times x_t + U_i \times h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma_g(W_o \times x_t + U_o \times h_{t-1} + b_o) \quad (3)$$

$$c'_t = \sigma_c(W_c \times x_t + U_c \times h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t \quad (5)$$

$$h_t = o_t \cdot \sigma_c(c_t) \quad (6)$$

where

$x(t)$ —LSTM unit input at time step;

$o(t)$ —LSTM unit output at time step;

$W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c$ —weight matrices;

b_f, b_i, b_o, b_c —biases;
 σ_s —sigmoid function;
 σ_c —tanh function;
 f_t —forget LSTM gate;
 i_t —input LSTM gate;
 o_t —output LSTM gate;
 c_t —cell state LSTM gate;
 h_t —hidden LSTM gate.

The models proposed in this study were developed using the Keras framework, and computations were performed using a CUDA graphic processing unit (Geforce 1060 with 6 GB of memory). The LSTM model has two neural network layers. The first one is the LSTM layer, and the second one is the dense layer. The network structure is shown in Table 3 and Figure 2. The network parameters of the LSTM model were chosen to allow for computations on a mid-range computer in a reasonable time.

Table 3. Model layers parameters.

Layer Type	Number of Units	Number of Params
LSTM	128	66,560
Danse	1	129
Total params		66,689

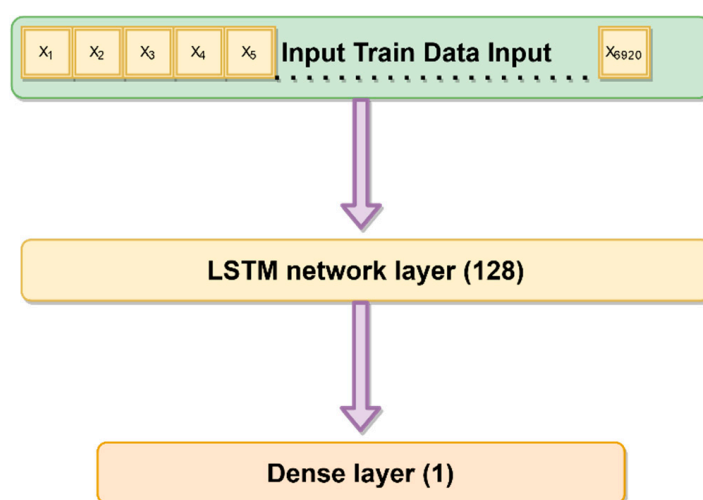


Figure 2. LSTM model network architecture.

The number of epochs was selected as 50, and increasing that value did not significantly enhance the learning process. The Adam optimiser algorithm was chosen, allowing for the first-order gradient-based optimisation of stochastic objective functions, described by Kingma et al. [57], with a 0.0001 value. The dataset was split into 80 per cent of data for training and 20 per cent for testing.

Before training, the dataset needed to be rescaled by so-called *robust data scaling*, provided by the scikit-learn library, as described by Pedregosa et al. [58]. The robust transformation is based on the median and interquartile range rather than the mean and standard deviation. Equation (7) shows how each value was scaled before the learning process.

$$\text{Value} = (\text{value} - \text{median}) / (\text{Q75} - \text{Q25}) \quad (7)$$

3.2. Deep Learning Process

The learning process described in this study is shown in Figure 3. The first step was to load the measurement data, e.g., in the .csv file format. The next step was preprocessing

and splitting data into training and test sets. After that, the model was created by choosing different values for the number of units in the LSTM layer and different prediction time steps. The last step was evaluating the model by comparing data from the prediction with the test set.

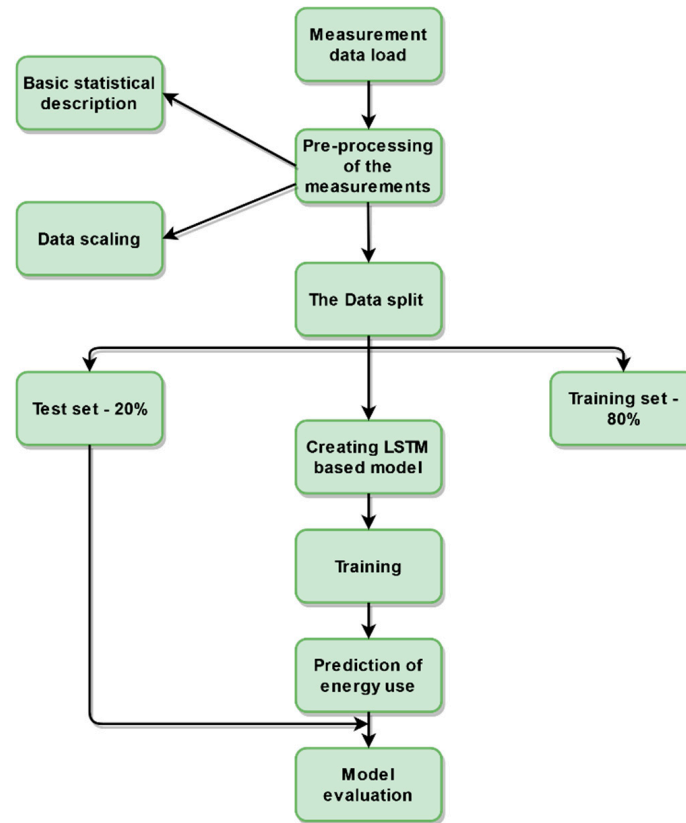


Figure 3. Learning process diagram.

4. Results

The results of different forecasting lengths are shown in Table 4 (1–3). The model learning loss plot is shown in Figure 4. Figure 5 shows the similarity of true (test data) and predicted (model result) values. The learning process took from 350 to 500 s.

Table 4. Comparison of prediction evaluation metrics for performed forecasts.

	Parameter	Prediction 1 h (360 Points)	Prediction 2 h (720 Points)	Prediction 3 h (1080 Points)	Prediction 4 h (1440 Points)
1	Single-layer LSTM mean-squared error	0.0112	0.0087	0.0084	0.0067
2	Single-layer LSTM mean absolute error	0.0524	0.0464	0.0487	0.0476
3	Single-layer LSTM cosine similarity	0.9039	0.9080	0.9195	0.9105
4	Double-layer LSTM mean-squared error	0.0119	0.0102	0.0111	0.0085
5	Double-layer LSTM mean absolute error	0.0714	0.0486	0.0494	0.0389
6	Double-layer LSTM cosine similarity	0.8565	0.9467	0.9588	0.9378
7	CNN network mean-squared error	0.0178	0.0497	0.1420	0.0885

Table 4. Cont.

	Parameter	Prediction 1 h (360 Points)	Prediction 2 h (720 Points)	Prediction 3 h (1080 Points)	Prediction 4 h (1440 Points)
8	CNN network mean absolute error	0.0844	0.1737	0.3272	0.2322
9	CNN network cosine similarity	0.8504	0.2058	−0.32133	0.5397

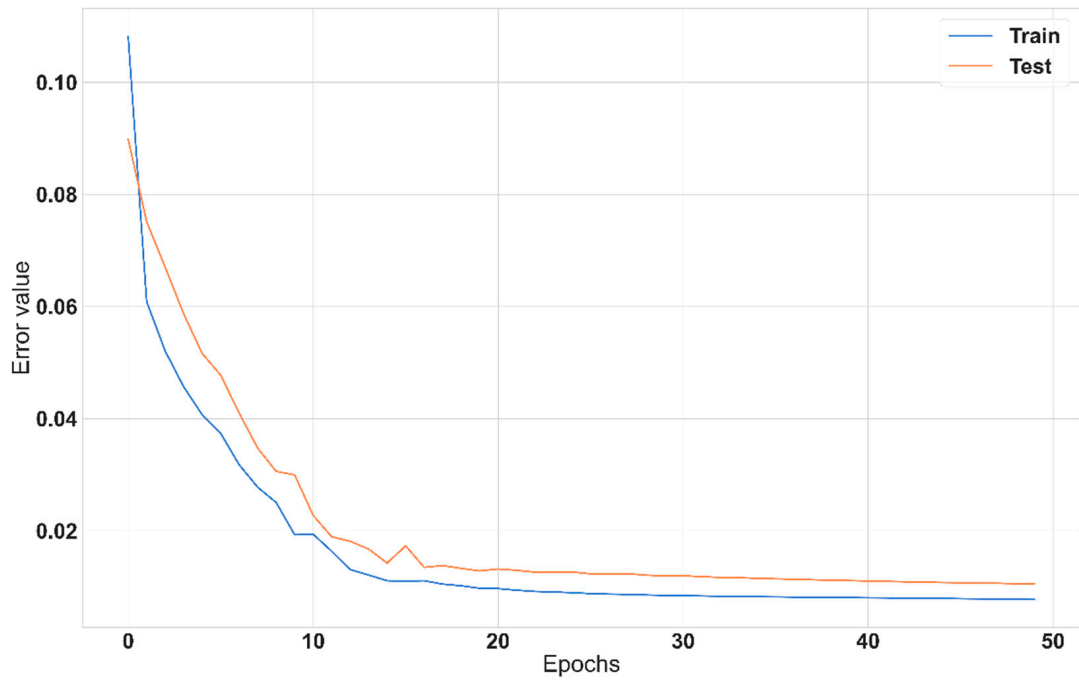


Figure 4. Training and testing loss-value error.

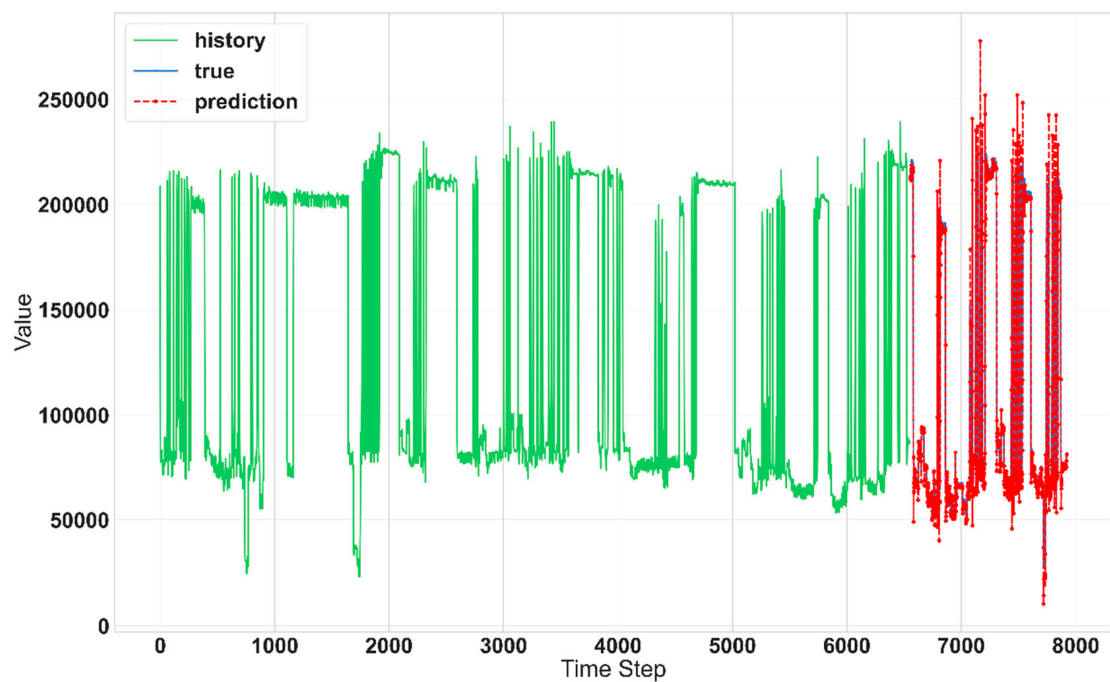


Figure 5. Actual and predicted data: green—training data (history), blue—test data (true), dotted red—prediction.

The proposed method was compared with two alternate network processes. The first is a double-layered LSTM network—one layer of 128 units and another of 64 units. Table 4 (4–6) shows the results. The second type of network is based on CNN architecture. The convolution layer comprises 32 filters of kernel size equal to 3 with the activation function *relu*. Then, there is a MaxPooling layer, with a pool size equal to two. Next, there is the flattened and dense layer. The mean-squared error metric as a loss function was used for learning the *Adam* optimiser. Table 4 (7–9) presents the results of the prediction metrics. Among them, mean-squared error (*MSE*), mean absolute error (*MAE*), and cosine similarity (*CS*) were used, which are defined in Equations (8)–(10). The learning graphs of all three network types are shown in Figures 6–9. All networks described in this study were created with Keras and TensorFlow frameworks.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (|Y_i - \hat{Y}_i|) \quad (9)$$

$$CS = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (10)$$

Learning for prediction of 360 data points - 1 hour

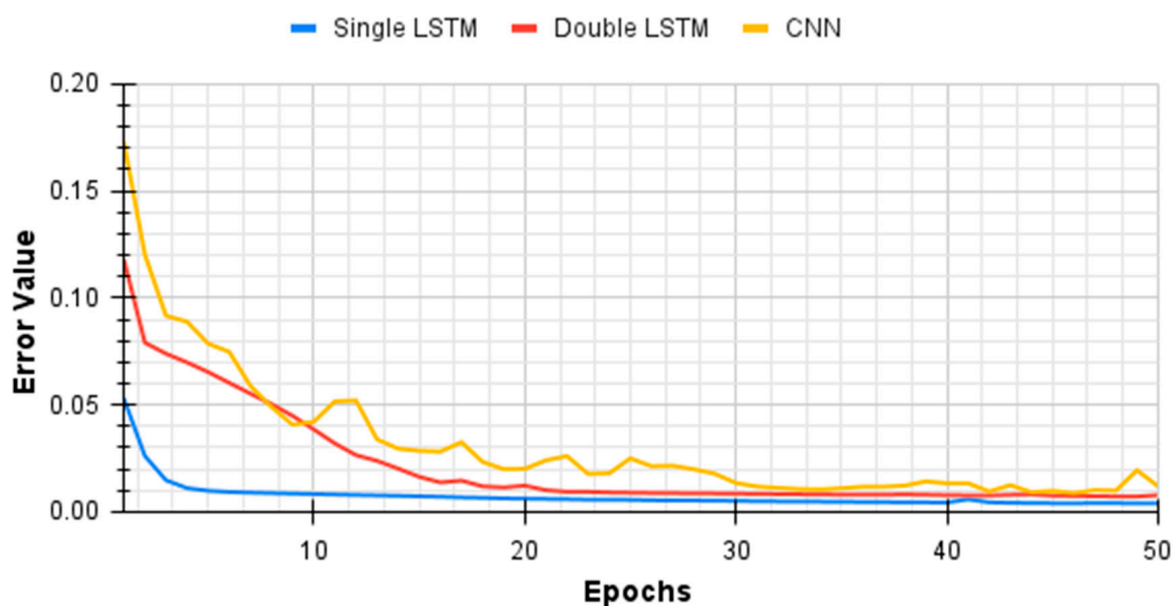


Figure 6. Comparison of the learning processes of different types of LSTM and CNN networks in the 1 h forecasting horizon: X-axis—next epochs, Y-axis—error level.

Results for single-layer LSTM gathered in Table 4 show a decrease in error values for *MSE* and *MAE* metrics with an increase in prediction horizon (rows 1 and 2). The same result is shown in Figures 6–9, when, after the decrease in error values, in the first few epochs, the level is constant. Cosine similarity (row 3) error value stays between 0.9039 and 0.9105 despite its 4 times longer prediction. Rows 4 and 5 show similar behaviour—with the difference being that the decrease in error value takes longer (up to 20 epochs). Cosine similarity is more fluctuated (row 6). The CNN network (rows 7–9) has the most fluctuating behaviour among error value plots. Figures 8 and 9 show peaks of error value increase. The longer the forecasting time, the more changeability can be observed. Another disadvantage

of the CNN network in comparison with single-layer LSTM and double-layer LSTM is the trend of increased error values with an increase in computation epochs.

Learning for prediction of 720 data points - 2 hours

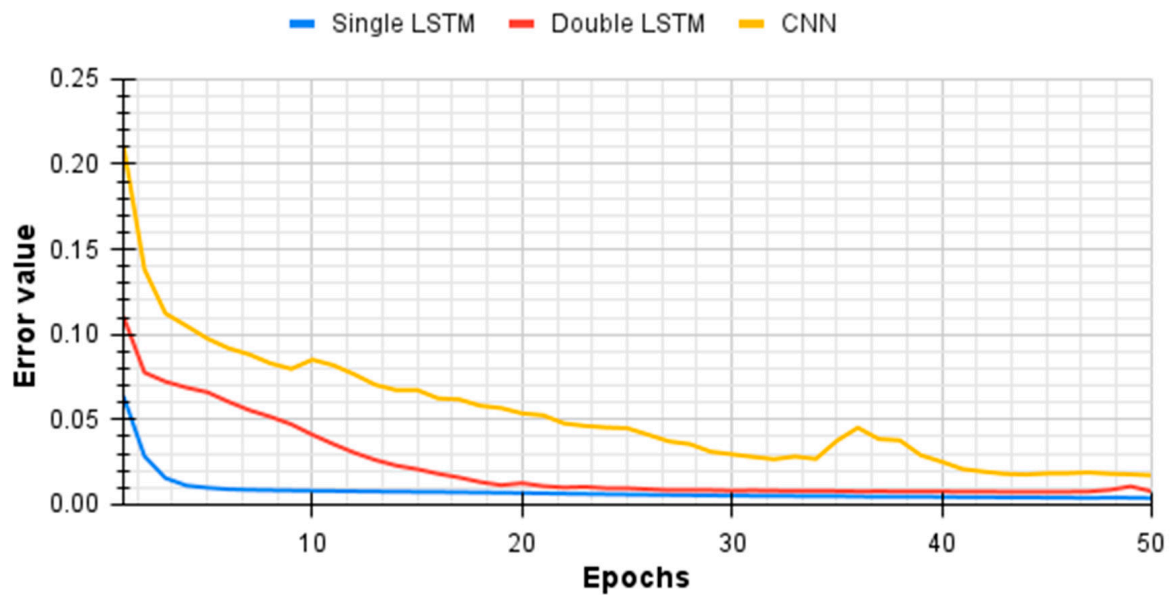


Figure 7. Comparison of the learning process of different types of LSTM and CNN networks in the 2 h forecasting horizon: X-axis—next epochs, Y-axis—error level.

Learning for prediction of 1080 data points - 3 hours

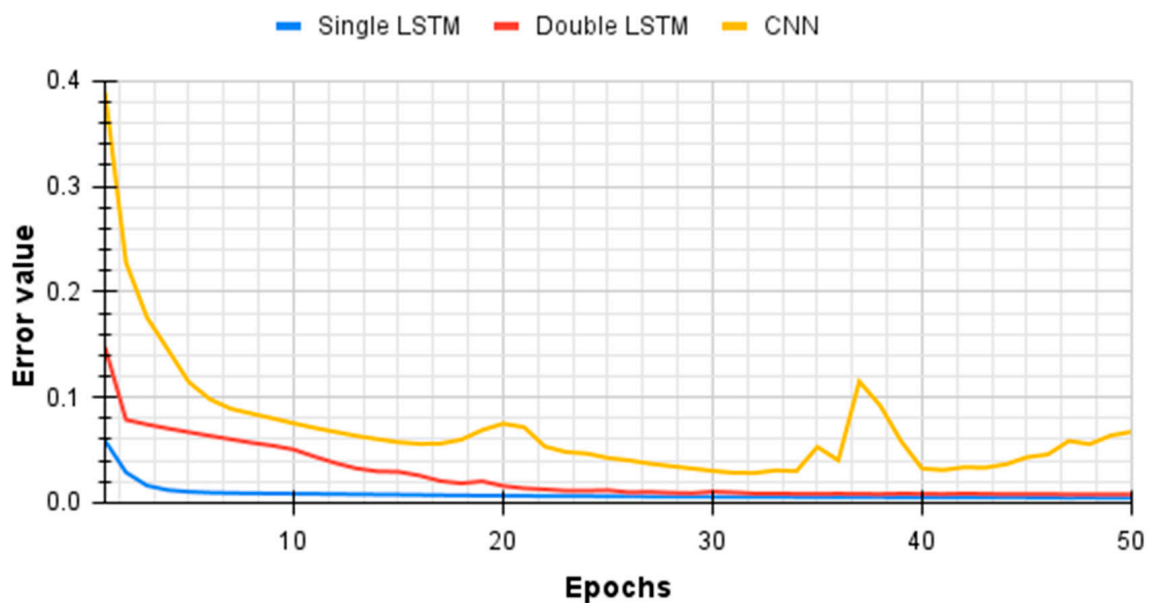


Figure 8. Comparison of the learning process of different types of LSTM and CNN networks in the 3 h forecasting horizon: X-axis—next epochs, Y-axis—error level.

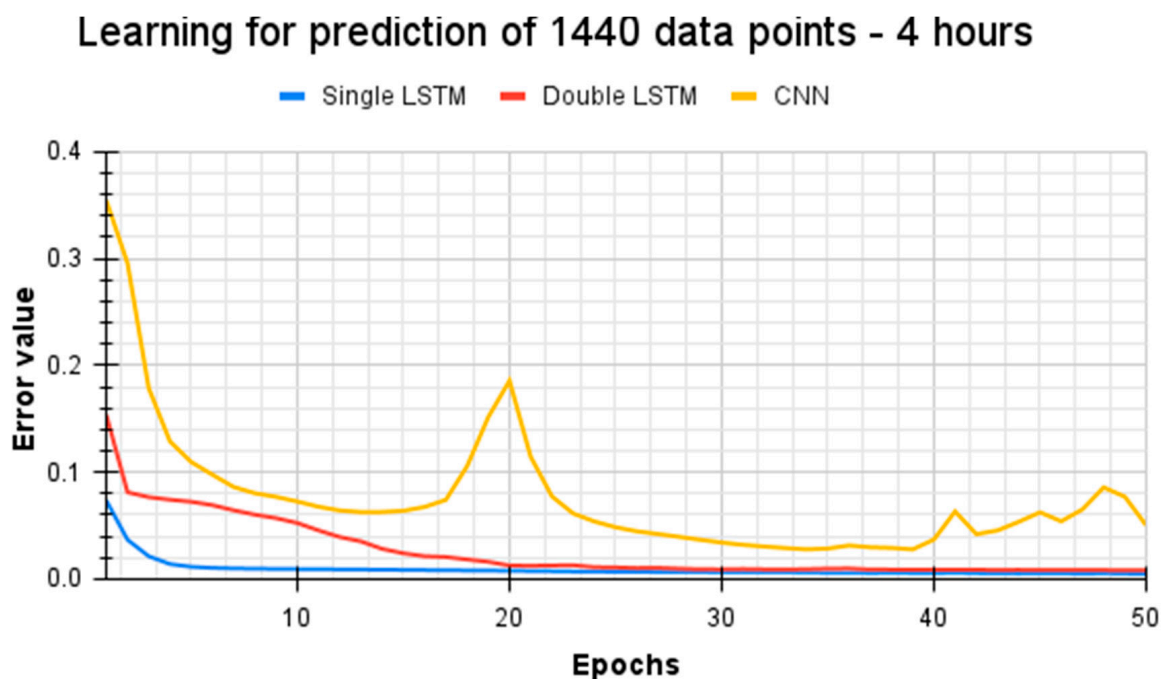


Figure 9. Comparison of the learning process of different types of LSTM and CNN networks in the 4 h forecasting horizon: X-axis—next epochs, Y-axis—error level.

Figures 6–9 present the properties of the single LSTM network as the model with the fastest converging learning rate (blue). In contrast, the model with double LSTM layers (red) has a slower learning rate, vulnerable to overfitting, and requires more memory resources; the CNN layer (yellow) is sensitive to the length of the data forecast. Furthermore, Figures 8 and 9 demonstrate that the longer the data prediction, the higher the instability, error, and learning rate. The CNN network is vulnerable to overfitting, creating a large scatter of results. Thus, the single-layer LSTM network shows the best convergence in the learning rate, which means the fastest minimisation of the error value.

Additional comparison of forecasts accuracy was carried out with the Diebold–Mariano (DM) test [59,60] application. In Table 5, results (p -values) of forecasting accuracy are presented for the shortest (1 h) forecast. Values in the table (Table 5) confirm better accuracy of the single-layer LSTM model in comparison to double-layer LSTM and CNN models. In fact, both LSTM models achieved better accuracy than the CNN model. As DM test results for other forecasting periods yield values very close to 0 or 1, the presentation of these data was considered as not contributing significantly to the discussed issue. However, it affirms what is observed in the case of 1 h forecasts. The obtained undifferentiated results of the DM test for longer forecasts (2 h and more) are possibly caused by the immanent qualities of the DM test in these types of data.

Table 5. Results of the Diebold–Mariano tests for 1 h forecasting accuracy with the null hypothesis that the model in the row predicts more precisely than the model in the column.

Models	Single-Layer LSTM	Double-Layer LSTM	CNN Layers
Single-layer LSTM	-	0.70	0.99
Double-Layer LSTM	0.30	-	0.99
CNN layers	<0.0001	<0.0001	-

5. Discussion

This study aimed to develop a useful model to forecast energy consumption in a short-term future window (0–4 h) based on previous energy use data. The proposed model

can predict energy demand based on only one measurement—energy consumption. The selection of LSTM as an ANN layer proved to be optimal for the above task and is a good compromise between the simplicity of network architecture and computation time. As shown in Figures 6–9, the proposed model with a single LSTM layer rapidly converges and maintains a constant and low value of error. This is especially important due to fact that forecasting energy consumption is measured at 10 s intervals. This feature of our model highlights its potential in application areas in the industry where there are rapidly changing values of measurement data.

In comparison with developed solutions, the more complicated double-layer LSTM model requires more epochs to successfully converge and minimise its error values. The CNN model shows the greatest instability and peaks of error values. Moreover, in the case of highly nonlinear data or when convergence time will not be a priority, our model can be expanded by adding new LSTM layers. Another advantage of the proposed model is its ability to be integrated into the so-called forecasting machine learning pipeline for constantly forecasting new prognoses. Due to the development of the model with industry-standard software—namely, the Keras framework and Python language—instead of academic-only statistical software, our proposed model can be easily integrated into other machine learning applications. The Diebold–Mariano test was used to compare forecasting accuracy.

The energy consumption forecasting method proposed in this study is feasible to adapt to an industry environment without specialised forecasting knowledge. Only the measurements of energy consumption data are required. Next to the economic benefits associated with the ability to predict power demand, it will allow manufacturers to become active players in the microgrid energy markets. Predicting energy consumption makes it possible to optimise the purchase, sale, and use of energy, and it provides flexibility and increases resilience to supply disruptions. The following systems are architectures for automated energy market design in smart grids and microgrids. Electricity consumption will be constantly increasing and access to the proposed tool allows prosumers (consumers with the ability to sell energy) to optimise financial gains.

Currently, new studies in the area of energy forecasting are focused on increasing the accuracy of predictions using LSTM networks along with other types of algorithms—the so-called hybrids methods. Extension of LSTM by connecting it with singular spectrum analysis was propped by Jin et al. [61]. LSTM network was connected with a CNN network to increase the ability of precise prediction by Agga et al. [62]. Another hybrid model was developed by the connection of random forests and LSTM algorithms layers [63].

Ultimately, it should be stated that the forecasting model based on the single-layer LSTM network proposed by the authors shows an acceptable compromise between accuracy and computation time and requirements. The authors emphasise industrial application possibilities related to the use of this forecasting model, e.g., for the prediction of a factory's energy consumption, with up to 4 h horizon.

6. Conclusions

From the perspective of end-users, energy consumption forecast is the barrier that limits the adoption and use of smart grids. In this study, the authors developed a model for short-term (0–4 h) energy demand prediction based on a manufacturing plant's measurement data. The presented model is adaptable to different manufacturing companies; it requires only previous energy use data. According to the data gathered in Table 3, a single-layer LSTM model can predict energy utilisation, with a mean absolute error value of 0.0464, which shows good model prediction capabilities. Additional visual inspection of Figure 5 shows a lack of differences between real and predicted values.

The single-layer LSTM model was also compared with two other approaches: the first was the double-layer LSTM, and the second was the CNN-based model. The double-layer LSTM has a slower convergence rate and higher memory and GPU time requirements due to more LSTM units. The CNN approach shows a lack of stability and the need

for a larger dataset which can be unfeasible in real manufacturing scenarios. The model presented by the authors allows end-users to optimise energy utilisation costs by delivering forecasts of energy use. In environments with limited power line throughput, it can be used as a tool for optimisation of machine load and scheduling production plans according to energy utilisation requirements of factory equipment. In future research, forecasting apparent and reactive power will be considered and our model will be expanded to include weather parameters. The authors plan to use hybrid or ensemble models to increase the accuracy of the forecasting model, e.g., by using larger datasets with more features and hyperparameter optimisation.

Author Contributions: Conceptualisation, W.U.; methodology, W.U. and M.S.; validation, W.U. and M.S.; formal analysis, W.U. and M.S.; investigation, W.U. and M.S., data curation, W.U. and M.S.; writing—original draft preparation, W.U. and M.S.; writing—review and editing, W.U. and M.S.; visualisation, M.S.; coding, M.S., supervision, W.U.; project administration, W.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant from the Minister of Science and Higher Education received by the Bialystok University of Technology, Grant Number W/WIZ/3/2022.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Considine, T.; Cox, W.; Cazalet, E.G. Understanding Microgrids as the Essential Architecture of Smart Energy. In Proceedings of the Grid-Interop Forum 2012, Irving, TX, USA, 3–6 December 2012. [CrossRef]
2. Ma, J.; Ma, X. A Review of Forecasting Algorithms and Energy Management Strategies for Microgrids. *Syst. Sci. Control. Eng.* **2018**, *6*, 237–248. [CrossRef]
3. Worighi, I.; Maach, A.; Hafid, A.; Hegazy, O.; Van Mierlo, J. Integrating Renewable Energy in Smart Grid System: Architecture, Virtualization and Analysis. *Sustain. Energy Grids Netw.* **2019**, *18*, 100226. [CrossRef]
4. Sabzehgar, R.; Amirhosseini, D.Z.; Rasouli, M. Solar Power Forecast for a Residential Smart Microgrid Based on Numerical Weather Predictions Using Artificial Intelligence Methods. *J. Build. Eng.* **2020**, *32*, 101629. [CrossRef]
5. Alotaibi, I.; Abido, M.A.; Khalid, M.; Savkin, A.V. A Comprehensive Review of Recent Advances in Smart Grids: A Sustainable Future with Renewable Energy Resources. *Energies* **2020**, *13*, 6269. [CrossRef]
6. Shulyma, O.; Shendryk, V.; Baranova, I.; Marchenko, A. The Features of the Smart MicroGrid as the Object of Information Modeling. In *Information and Software Technologies*, 2nd ed.; Dregvaite, G., Damasevicius, R., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2014; Volume 465, pp. 12–23. [CrossRef]
7. Palma-Behnke, R.; Reyes, L.; Jimenez-Estevéz, G. Smart Grid Solutions for Rural Areas. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012; pp. 1–6. [CrossRef]
8. Masembe, A. Reliability Benefit of Smart Grid Technologies: A Case for South Africa. *J. Energy S. Afr.* **2015**, *26*, 2–9. [CrossRef]
9. Samad, T.; Kiliccote, S. Smart Grid Technologies and Applications for the Industrial Sector. *Comput. Chem. Eng.* **2012**, *47*, 76–84. [CrossRef]
10. Keller, F.; Schultz, C.; Simon, P.; Braunreuther, S.; Glasschröder, J.; Reinhart, G. Integration and Interaction of Energy Flexible Manufacturing Systems within a Smart Grid. *Procedia CIRP* **2017**, *61*, 416–421. [CrossRef]
11. Shahid, A. Smart Grid Integration of Renewable Energy Systems. In Proceedings of the 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, 14–17 October 2018; pp. 944–948. [CrossRef]
12. Yaprakdal, F.; Yilmaz, M.B.; Baysal, M.; Anvari-Moghaddam, A. A Deep Neural Network-Assisted Approach to Enhance Short-Term Optimal Operational Scheduling of a Microgrid. *Sustainability* **2020**, *12*, 1653. [CrossRef]
13. Elattar, E.E.; Sabiha, N.A.; Alsharif, M.; Metwaly, M.K.; Abd-Elhady, A.M.; Taha, I.B.M. Short Term Electric Load Forecasting Using Hybrid Algorithm for Smart Cities. *Appl. Intell.* **2020**, *50*, 3379–3399. [CrossRef]
14. Wood, D.A. Hourly-Averaged Solar plus Wind Power Generation for Germany 2016: Long-Term Prediction, Short-Term Forecasting, Data Mining and Outlier Analysis. *Sustain. Cities Soc.* **2020**, *60*, 102227. [CrossRef]
15. Ahmad, A.; Javaid, N.; Mateen, A.; Awais, M.; Khan, Z.A. Short-Term Load Forecasting in Smart Grids: An Intelligent Modular Approach. *Energies* **2019**, *12*, 164. [CrossRef]
16. Wang, Y.; Huang, Y.; Wang, Y.; Li, F.; Zhang, Y.; Tian, C. Operation Optimization in a Smart Micro-Grid in the Presence of Distributed Generation and Demand Response. *Sustainability* **2018**, *10*, 847. [CrossRef]
17. Kempener, R.; Komor, P.; Hoke, A. Smart Grids and Renewables, A Guide for Effective Deployment, Working Paper. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2013/smart_grids.pdf?la=en&hash=08F3E571B5580F017E70BCD1EC39864536681ADB (accessed on 8 October 2021).

18. Siami-Namini, S.; Tavakoli, N.; Siami Namin, A. A Comparison of ARIMA and LSTM in Forecasting Time Series. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1394–1401. [[CrossRef](#)]
19. Yang, H.-T.; Huang, C.-M.; Huang, C.-L. Identification of ARMAX Model for Short Term Load Forecasting: An Evolutionary Programming Approach. *IEEE Trans. Power Syst.* **1996**, *11*, 403–408. [[CrossRef](#)]
20. Fang, T.; Lahdelma, R. Evaluation of a Multiple Linear Regression Model and SARIMA Model in Forecasting Heat Demand for District Heating System. *Appl. Energy* **2016**, *179*, 544–552. [[CrossRef](#)]
21. Su, W.; Wang, J.; Zhang, K.; Huang, A.Q. Model Predictive Control-Based Power Dispatch for Distribution System Considering Plug-in Electric Vehicle Uncertainty. *Electr. Power Syst. Res.* **2014**, *106*, 29–35. [[CrossRef](#)]
22. Emami, A.; Sarvi, M.; Asadi Bagloee, S. Using Kalman Filter Algorithm for Short-Term Traffic Flow Prediction in a Connected Vehicle Environment. *J. Mod. Transport.* **2019**, *27*, 222–232. [[CrossRef](#)]
23. Shah, I.; Bibi, H.; Ali, S.; Wang, L.; Yue, Z. Forecasting One-Day-Ahead Electricity Prices for Italian Electricity Market Using Parametric and Nonparametric Approaches. *IEEE Access* **2020**, *8*, 123104–123113. [[CrossRef](#)]
24. Aykroyd, G.R.; Alfaer, N. Sequential Models for Time-evolving Regression Problems with an Application to Energy Demand Prediction. *Stoch. Modeling Appl.* **2016**, *20*, 1–16.
25. Lisi, F.; Shah, I. Forecasting Next-Day Electricity Demand and Prices Based on Functional Models. *Energy Syst.* **2020**, *11*, 947–979. [[CrossRef](#)]
26. Hirose, K.; Wada, K.; Hori, M.; Taniguchi, R. Event Effects Estimation on Electricity Demand Forecasting. *Energies* **2020**, *13*, 5839. [[CrossRef](#)]
27. Bibi, N.; Shah, I.; Alsubie, A.; Ali, S.; Lone, S.A. Electricity Spot Prices Forecasting Based on Ensemble Learning. *IEEE Access* **2021**, *9*, 150984–150992. [[CrossRef](#)]
28. Shah, I.; Iftikhar, H.; Ali, S. Modeling and Forecasting Medium-Term Electricity Consumption Using Component Estimation Technique. *Forecasting* **2020**, *2*, 163–179. [[CrossRef](#)]
29. Zhang, W.; Zhang, H.; Liu, J.; Li, K.; Yang, D.; Tian, H. Weather Prediction with Multiclass Support Vector Machines in the Fault Detection of Photovoltaic System. *IEEE/CAA J. Autom. Sinica* **2017**, *4*, 520–525. [[CrossRef](#)]
30. del Real, A.J.; Dorado, F.; Durán, J. Energy Demand Forecasting Using Deep Learning: Applications for the French Grid. *Energies* **2020**, *13*, 2242. [[CrossRef](#)]
31. Kang, T.; Lim, D.Y.; Tayara, H.; Chong, K.T. Forecasting of Power Demands Using Deep Learning. *Appl. Sci.* **2020**, *10*, 7241. [[CrossRef](#)]
32. Kumar, S.; Hussain, L.; Banarjee, S.; Reza, M. Energy Load Forecasting Using Deep Learning Approach-LSTM and GRU in Spark Cluster. In Proceedings of the 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 12–13 January 2018; pp. 1–4. [[CrossRef](#)]
33. Mele, E.; Elias, C.; Ktena, A. Machine Learning Platform for Profiling and Forecasting at Microgrid Level. *Electr. Control. Commun. Eng.* **2019**, *15*, 21–29. [[CrossRef](#)]
34. Konstantinou, M.; Peratikou, S.; Charalambides, A.G. Solar Photovoltaic Forecasting of Power Output Using LSTM Networks. *Atmosphere* **2021**, *12*, 124. [[CrossRef](#)]
35. Ahn, H.K.; Park, N. Deep RNN-Based Photovoltaic Power Short-Term Forecast Using Power IoT Sensors. *Energies* **2021**, *14*, 436. [[CrossRef](#)]
36. Brahma, B.; Wadhvani, R. Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data. *Symmetry* **2020**, *12*, 1830. [[CrossRef](#)]
37. Jeon, B.; Kim, E.-J. Next-Day Prediction of Hourly Solar Irradiance Using Local Weather Forecasts and LSTM Trained with Non-Local Data. *Energies* **2020**, *13*, 5258. [[CrossRef](#)]
38. Bilgili, M.; Arslan, N.; Sekertekin, A.; Yasar, A. Application of long short-term memory (LSTM) neural network based on deep learning for electricity energy consumption forecasting. *Turk. J. Elec. Eng. Comp. Sci.* **2022**, *30*, 140–157. [[CrossRef](#)]
39. Peng, L.; Wang, L.; Xia, D.; Gao, Q. Effective Energy Consumption Forecasting Using Empirical Wavelet Transform and Long Short-Term Memory. *Energy* **2022**, *238*, 121756. [[CrossRef](#)]
40. Luo, X.J.; Oyedele, L.O. Forecasting Building Energy Consumption: Adaptive Long-Short Term Memory Neural Networks Driven by Genetic Algorithm. *Adv. Eng. Inform.* **2021**, *50*, 101357. [[CrossRef](#)]
41. Pena-Gallardo, R.; Medina-Rios, A. A Comparison of Deep Learning Methods for Wind Speed Forecasting. In Proceedings of the 2020 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 4–6 November 2020; pp. 1–6. [[CrossRef](#)]
42. Pao, H. Comparing Linear and Nonlinear Forecasts for Taiwan’s Electricity Consumption. *Energy* **2006**, *31*, 2129–2141. [[CrossRef](#)]
43. Wang, J.Q.; Du, Y.; Wang, J. LSTM Based Long-Term Energy Consumption Prediction with Periodicity. *Energy* **2020**, *197*, 117197. [[CrossRef](#)]
44. Laib, O.; Khadir, M.T.; Mihaylova, L. Toward Efficient Energy Systems Based on Natural Gas Consumption Prediction with LSTM Recurrent Neural Networks. *Energy* **2019**, *177*, 530–542. [[CrossRef](#)]
45. Deligiannidis, S.; Mesaritakis, C.; Bogris, A. Performance and Complexity Evaluation of Recurrent Neural Network Models for Fibre Nonlinear Equalization in Digital Coherent Systems. In *2020 European Conference on Optical Communications (ECOC)*; IEEE: Brussels, Belgium, 2020; pp. 1–4. [[CrossRef](#)]

46. Kaur, D.; Islam, S.N.; Mahmud, M.A.; Dong, Z. Energy Forecasting in Smart Grid Systems: A Review of the State-of-the-Art Techniques. *arXiv* **2020**, arXiv:2011.12598.
47. Choi, J.Y.; Lee, B. Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting. *Math. Probl. Eng.* **2018**, *2018*, 2470171. [[CrossRef](#)]
48. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [[CrossRef](#)]
49. Sun, Y.; Haghighat, F.; Fung, B.C.M. A Review of The-State-of-the-Art in Data-Driven Approaches for Building Energy Prediction. *Energy Build.* **2020**, *221*, 110022. [[CrossRef](#)]
50. Somu, N.; Raman, M.R.G.; Ramamritham, K. A Deep Learning Framework for Building Energy Consumption Forecast. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110591. [[CrossRef](#)]
51. Luo, X.; Zhang, D.; Zhu, X. Deep Learning Based Forecasting of Photovoltaic Power Generation by Incorporating Domain Knowledge. *Energy* **2021**, *225*, 120240. [[CrossRef](#)]
52. Manowska, A. Using the LSTM Network to Forecast the Demand for Electricity in Poland. *Appl. Sci.* **2020**, *10*, 8455. [[CrossRef](#)]
53. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
54. Graves, A.; Liwicki, M.; Fernandez, S.; Bertolami, R.; Bunke, H.; Schmidhuber, J. A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 855–868. [[CrossRef](#)]
55. Sak, H.; Senior, A.; Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *arXiv* **2014**, arXiv:1402.1128.
56. Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv* **2016**, arXiv:1609.08144.
57. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
58. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [[CrossRef](#)]
59. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [[CrossRef](#)]
60. Shah, I.; Iftikhar, H.; Ali, S.; Wang, D. Short-term electricity demand forecasting using components estimation technique. *Energies* **2019**, *12*, 2532. [[CrossRef](#)]
61. Jin, N.; Yang, F.; Mo, Y.; Zeng, Y.; Zhou, X.; Yan, K.; Ma, X. Highly Accurate Energy Consumption Forecasting Model Based on Parallel LSTM Neural Networks. *Adv. Eng. Inform.* **2022**, *51*, 101442. [[CrossRef](#)]
62. Agga, A.; Abbou, A.; Labbadi, M.; Houm, Y.E.; Ou Ali, I.H. CNN-LSTM: An Efficient Hybrid Deep Learning Architecture for Predicting Short-Term Photovoltaic Power Production. *Electr. Power Syst. Res.* **2022**, *208*, 107908. [[CrossRef](#)]
63. Karijadi, I.; Chou, S.-Y. A Hybrid RF-LSTM Based on CEEMDAN for Improving the Accuracy of Building Energy Consumption Prediction. *Energy Build.* **2022**, *259*, 111908. [[CrossRef](#)]