energies

MDPI

Article

Long-Term Electricity Demand Forecasting in the Steel Complex Micro-Grid Electricity Supply Chain—A Coupled Approach

Sepehr Moalem [1], Roya M. Ahari [1], Ghazanfar Shahgholian [2,3,*], Majid Moazzami [2,3] and Seyed Mohammad Kazemi [1]

[1] Industrial Engineering Department, Najafabad Branch, Islamic Azad University, Najafabad 8514143131, Iran
[2] Department of Electrical Engineering, Najafabad Branch, Islamic Azad University, Najafabad 8514143131, Iran
[3] Smart Microgrid Research Center, Najafabad Branch, Islamic Azad University, Najafabad 8514143131, Iran
* Correspondence: shahgholian@iaun.ac.ir

Abstract: Demand forecasting produces valuable information for optimal supply chain management. The basic metals industry is the most energy-intensive industries in the electricity supply chain. There are some differences between this chain and other supply chains including the impossibility of large-scale energy storage, reservation constraints, high costs, limitations on electricity transmission lines capacity, real-time response to high-priority strategic demand, and a variety of energy rates at different hours and seasons. A coupled demand forecasting approach is presented in this paper to forecast the demand time series of the metal industries microgrid with minimum available input data (only demand time series). The proposed method consists of wavelet decomposition in the first step. The training subsets and the validation subsets are used in the training and fine-tuning of the LSTM model using the ELATLBO method. The ESC dataset used in this study for electrical demand forecasting includes 24-h daily over 40 months from 21 March 2017, to 21 June 2020. The obtained results have been compared with the results of Support Vector Machine (SVM), Decision Tree, Boosted Tree, and Random Forest forecasting models optimized using the Bayesian Optimization (BO) method. The results show that performance of the proposed method is well in demand forecasting of the metal industries.

Keywords: electricity supply chain; long short-term memory neural network (LSTM); hyper-parameter; ELATLBO; wavelet transform; micro-grid; Bayesian optimization

check for updates

Citation: Moalem, S.; Ahari, R.M.; Shahgholian, G.; Moazzami, M.; Kazemi, S.M. Long-Term Electricity Demand Forecasting in the Steel Complex Micro-Grid Electricity Supply Chain—A Coupled Approach. *Energies* **2022**, *15*, 7972. https://doi.org/10.3390/en15217972

Academic Editor: Tek Tjing Lie

Received: 12 September 2022
Accepted: 24 October 2022
Published: 27 October 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

1.1. Background and Motivation

Forecasting energy consumption is crucial for the designation of an energy supply chain. The electricity supply chain is a large-scale energy supply network with a wide range of consumers and it must have the necessary ability to respond to the requirements of the applicants under this supply chain with the desired quality and speed.

Electricity energy must be consumed at the same time as it is generated in the power plant due to its expensive storage costs [1]. As a result, to keep the power system balanced at the levels of generation, transmission, and distribution, the balance of electricity generation and consumption must always be maintained. The need to maintain a power balance in the electricity supply chain depends on accurate demand forecasting.

Over estimation and under estimation of electricity demand each, in turn, lead to additional costs, which is undesirable in the electricity supply chain. Demand fluctuations and the inability to respond promptly to demand from the electricity supply chain will lead to disruption and cascade errors to other levels of the chain and ultimately lead to the collapse of the electricity supply chain. Lack of strategic and macro-planning and allocation of credit lines for infrastructure development, in addition to causing fluctuations and disruptions in the electricity supply chain, increases backward costs and frequent

disruptions in the production and service chain of related industries and it depends on this chain.

Large consumers play an important role in determining total electricity demand in the electricity supply chain. Among large consumers in Iran, the metal industry has the largest consumers in the electricity supply chain.

Hyper-parameter adjustment is an important step in training forecasting models with machine learning which can be done manually or automatically using certain methods. The technique of retrieving hyper-parameters requires tremendous computational resources, so it is almost impossible to reproduce experiments [2].

The main idea of this research is to automatically adjust hyper parameters of demand forecasting models to reduce complexity of manual hyperparameters adjustment. Hyper-parameter optimization of the forecasting model aims to achieve optimal forecasting results with the lowest possible errors.

1.2. Literature Review

Given the depletion of natural resources and fossil fuels, environmental pollution, and water shortage crises, the possibility of pandemics such as COVID-19, the need for protection and optimal management of these resources for future generations and life on earth, especially in the discussion of energy supply chains is one of the particular importance. An optimization model for the natural gas supply chain at three levels of refining, transmission, and distribution stages optimizes resources and costs by integrating DEA and fuzzy techniques presented in [3,4] offers an analytical hierarchical process for sustainable development of supply chain by providing a framework for decision making and methodology. It is one of the studies that show researchers pay attention to the stability and improvement of the situation in energy supply chains.

A review of research on electrical demand forecasting in recent years shows that the major studies have been based on regression-based and computational intelligence techniques respectively. In terms of time horizon, forecasts are made in three time periods: short-term, medium-term, and long-term [1]. Due to less uncertainty and the availability of statistical sources, most of the techniques used have focused on short-term forecasting.

Many factors made long-term forecasting less attractive to researchers such as uncertainties and lack of reliable data. On the other hand, short-term and medium-term demand forecasting requires high processing speed for real-time decision-making processes. Therefore, short-term and medium-term decision-making has a higher priority compared to long-term demand forecast periods. Demand forecasting, in the long run, does not have processing speed limits for fast decision-making and is mainly programmatic and done offline. Propagation is the provision of credit lines and the expansion of the necessary infrastructure to be able to respond promptly to future demand to reduce costs in the long run.

Accurate and reliable forecasts of upcoming demand provide critical information for supply chain managers that sharing it at different levels improves the quality of planning, inventory, and support for decisions made. In [5], the performance and accuracy of forecasting models due to demand fluctuations in both advertising and non-advertising periods are discussed and statistical models based on regression such as ARIMA and techniques based on machine learning are examined. This study shows that the number of demand fluctuations can affect the forecasting performance and is an important factor in choosing forecasting models that leads to the production of strong forecasts. One of the most frequently used methods in demand forecasting is the support vector machine (SVM) [6]. This method has been used independently or in combination with optimization methods such as particle swarm optimization (PSO). PSO is a method that is widely used to obtain the optimal values of hyper-parameters of SVM and to train demand forecasting models. The least-squares SVM (LS-SVM) combined with PSO [7] and the obtained results proved the effectiveness of the proposed method in short-term demand forecasting [7]. The SVM method is also used for short-term demand forecasting in the electricity distribution

system and a forecasting approach for short-term demand forecasting with high-resolution using vectorized regression and optimization of the hyper-parameters presented in [8]. Based on the sloping surface test and orthogonal support of the support vector machine, a hybrid forecasting method for the power generated by wind energy has been performed. The effectiveness of the method was tested with data from three wind farms in China and finally compared the results with other methods are discussed [9].

Among the large sets of forecasting models available for short-term forecasting, quantile regression averaging has been dramatically developed and principal component analysis (PCA) utilized for the automatic selection of forecasting models [10]. Probabilistic demand forecasting provides more information about the variability and uncertainty of future demand, which is very important for the scheduling and operation of electricity generation and transmission systems. Quantile regression forecasting based on sister forecasts produced by multiple linear regression (MLR) has been proposed in [11]. A study was presented in [12] to improve the neural network based on quantile regression. In this study, probabilistic demand forecasts in different intervals are presented. The data set selected for the case study conducted in this paper is taken from the historical hourly demand data [12].

By using the weighted fuzzy model and the experimental state decomposition process and SVM optimization through the bat algorithm and using the Kalman filter, a combined approach for short-term demand forecasting has been proposed [13]. Numerical case studies on the demand forecast of a substation in southern China show that the proposed combined forecasting model performs better than other forecasting methods and effectively improves forecasting accuracy [13]. In [14] a demand forecasting method based on multi-purpose data and a day-to-day topological network is presented. In addition, the random walk algorithm is applied by restarting the previous topology network to construct the training set and teach the prediction model using the vector machine regression model [14].

A recurrent deep neural network model has been used for hourly demand forecasting [15] and an attempt has been made to control the overfit problem by increasing the number and variety of model input data. The proposed method on the Tensor flow platform has been developed and then tested on 920 customers equipped with an intelligent measurement system. Compared to the most advanced residential demand forecasting methods, the proposed method has a higher performance [15].

In [16], a short-term residential demand forecasting method based on the Long Short-Term Memory network (LSTM) is presented. The proposed method has been tested on a dataset including demand data extracted from smart meters and its performance has been compared with various demand forecasting methods. The results show that the proposed LSTM method performed well in comparison with other residential demand forecasting algorithms [16]. In another study, a neural network based on a modified method of the Levenberg-Marquardt algorithm was conducted to day-ahead electrical demand forecasting in the electricity market [17]. Based on the combination of a generalized extreme learning machine, wavelet neural network, and bootstrap a hybrid method was proposed for probabilistic demand forecasting [18] and tested using the Ontario dataset and Australian electricity market. The results show the high performance and accuracy of the proposed method [18].

A hybrid model based on a recurrent neural network and dynamic time method is proposed [19] to forecast the daily electricity peak demand. A bottom-up approach to spatial data was used in [20] to predict transient demand using kernel density estimation and adaptive averaging method for aggregate profile and demand density, and stacked auto-encoders were utilized to predict unspecified quantities of electrical demand. In another study, a data-driven approach based on deep learning methods was proposed [21] for short-term demand forecasting. In this study, the data are processed using the Box-Cox transform, and the correlation between parameters such as electricity price, temperature, and demand is analyzed using Copula parametric models and the calculation of peak demand thresholds has been analyzed. In [22] a method based on Residual Neural Network (RNN) is presented. This method includes a two-step strategy to improve the ability of the

network for generalization. The Monte Carlo method has also been used to make possible predictions of electricity demand.

A multi dimensional convolution neural network (CNN) model with time cognition was proposed in [23] for short-term demand forecasting. In this method, a CNN model extracts the characteristics of different levels of data and the data are random and for 24 h a day. In [24], a neural network-based method is proposed for probabilistic net demand forecasting by considering the effect of solar power generation based on ISO New England data.

In finding the optimal hyper-parameters of forecasting models, the grid search method has been used by [25]. It skips the search space slightly for each iteration [25]. However, it is likely to skip between certain intervals, which is likely to produce high-performance values [26]. The random search used in [27] aims to find a global optimum. A Meta-heuristic optimization algorithm also can be used to adjust optimal hyper-parameters of forecasting models. The genetic algorithm was utilized to adjust the hyper-parameters of convolutional neural networks by evaluating the fitness of each hyper-parameter over the generations [28]. Particle Swarm Optimization (PSO) algorithm was recently utilized to train a recurrent neural network for voltage instability prediction. PSO algorithm applied in another study to find optimal hyper-parameters of CNN-LSTM forecasting model proposed for residential load forecasting [26] and it is also used to optimize weights of the LSTM model for stock forecasting [29].

The Differential evolution algorithm has been used by [30] to find optimal hyper-parameters of the LSTM model for forecasting the stock market. The hyper-parameters to optimize in this study are the size of the time window, batch size, the number of LSTM units in hidden layers, the number of hidden layers (LSTM and dense), the dropout coefficient for each layer, and the network training optimization algorithm [30].

In [31] a hybrid method proposed based on using Business Intelligence and Machine Learning. The machine learning engine executes data from different modules and determines the weekly, monthly, and quarterly demands of goods/commodities [31]. The purposed solution on real-time organization data get up to 92.38% accuracies for the store in terms of intelligent demand forecasting [31]. Deep boosting transfer regression proposed in [32] to forecast residential demand forecasting. A set of deep regression models on source houses firstly trained that can provide relatively abundant data. Then these learned models transfer via the boosting framework to support data-scarce target houses [32]. In [33], a hybrid algorithm that combines empirical mode decomposition (EMD), isometric mapping (Isomap), and Adaboost to construct a prediction mode for mid- to long-term load forecasting is developed [33]. A hybrid framework for short-term electricity demand forecasting proposed in [34] which includes data cleaning and a Residual Convolutional Neural Network (R-CNN) with multilayered Long Short-Term Memory (ML-LSTM) architecture. The proposed forecasting model in [35] comprises three modules. A pre-processing module; a forecast module, and an optimization module used in [35]. In the first module, correlated lagged load data along with influential meteorological and exogenous variables are fed as inputs to a feature selection technique which removes irrelevant and/or redundant samples from the inputs [35].

A deep learning restricted Boltzmann machine (RBM) is proposed in [36] for modelling and forecasting energy consumption. The contrastive divergence algorithm is presented for tuning the parameters and a statistical approach is suggested to determine the effective input variables [36]. In [37] three techniques are utilized for short-term load forecasting. These techniques are deep neural network (DNN), multilayer perceptron based artificial neural network (ANN), and decision tree-based prediction (DR) [37]. New predictive variables are included to enhance the overall forecasting and handle the difficulties caused by some categorical predictors [37]. An ensemble forecasting model proposed in [38] based on wavelet transform for the short-term load forecasting depending on the decomposition principle of load profiles. The model can effectively capture the portion of daily load profiles caused by seasonal variations [38]. The methodology proposed in [39] is based

on the long short-term memory (LSTM) model and penalized quantile regression (PQR). A comprehensive analysis for a time period of a year is conducted using the proposed method, and back propagation neural networks (BPNN), support vector machine (SVM), and random forest are applied as reference models [39]. In [40] an unsupervised multi-dimensional feature learning forecasting model proposed, named Multi DBN-T, based on a deep belief network and transformer encoder to accurately forecast short-term power load demand and implement power generation planning and scheduling [40]. In the model, the first layer (pre-DBN), based on a deep belief network, was designed to perform unsupervised multi-feature extraction feature learning on the data, and strongly coupled features between multiple independent observable variables were obtained [40].

The optimization algorithms used to tune the hyper-parameters of demand forecasting models have in turn many parameters to tune. Therefore, an optimization algorithm with a limited number of parameters to tune [40], which can perform as well as other optimization algorithms and produce even better results, will be the optimal solution for demand forecasting in the electricity supply chain.

1.3. Contributions and Paper Organization

This paper presents a coupled long-term approach to forecast electricity demand in metal industries micro-grid based on the combination of wavelets with an LSTM forecast engine optimized by an adaptive teaching-learning-based optimization with experience learning (ELATLBO). The under-study data sets in this paper are related to the Esfahan Steel Company (ESC) microgrid. This industry is not only one of the 10 most energy-consuming industries in the power supply chain of Isfahan province, but also it is one of the main suppliers of Iran's steel industry.

The proposed method in this study utilizes a combination of methods for accurate long-term demand forecasting of metal industries micro-grid. In the first part of the proposed method, wavelet analysis is used to produce predictor matrices which is the input for LSTM forecasting engines. In the second part, an enhanced teaching-learning-based optimization method will be used to optimally adjust the hyper-parameters of the LSTM forecasting engine. The data used in this study include 24-h daily over 40 months from 21 March 2017, to 21 June 2020. As a result, the only available data in this study is measured demand signal in terms of [MW].

The rest of the paper is organized as follows. Section 2 describes the proposed method and Section 3 describes the simulation results according to the proposed method and the diagrams and output results of each step. Finally, conclusions are presented in Section 4.

2. Methodology

The coupled approach framework consists of two main sections. The demand signal is decomposed in two signals using a low pass filter and a high pass filter applied to the main signal by means of a wavelet. A forecasting model will be trained using each signal then while optimizing the hyper-parameters of the forecasting model trough iterations of the optimization algorithm. Each signal is divided into three parts, which are training, validation and testing. A model will be trained using training part of each signal and LSTM with initial values of hyper-parameters at first. Then in validation section, the results will be produced using trained model and validation data and the RMSE will be calculated which is the objective function that should be optimized using ELATLBO. The design parameters in this optimization are hyper-parameters of LSTM model. In each iteration the LSTM will be trained using design parameter values and RMSE will be produced until the termination criteria ends the optimization. The LSTM model will be trained with the best values of hyper-parameters in this way that produce the least RMSE.

2.1. Discrete Wavelet Decomposition

Wavelets can be described as a pulse of short duration with finite energy that integrates to zero. The basic fact about them is that they are located in time (or space), unlike trigono-

metric functions [41]. This characteristic enables them to analyze many non-stationary signals [41]. Wavelet analysis utilizes mother wavelets. This function has a null mean and sharply drops in an oscillatory way. Data are represented via super position of scaled and translated versions of the pre-specified mother wavelet. The mother wavelet can be scaled and translated using certain scales and positions based on the powers of two [41]. This scheme is more efficient and just as accurate as the continuous wavelet transform (CWT). It is known as the discrete wavelet transform (DWT) and is defined as follows [41]:

$$\mathrm{DWT}(m,k) = \frac{1}{\sqrt{a_0^m}} \sum_n x(n) g\left(\frac{k - nb_0 a_0^m}{a_0^m}\right) \quad (1)$$

The scaling and translation parameters a and b in Equation (1) are functions of the integer variable $m(a = a_0^m \,\&\, b = nb_0a_0^m)$. The integer variable refers to a particular point of the input signal is k, and the discrete-time index is n. As the slope disappears, more data fitting occurs in the training phase, and such a model is not able to be generalized during the validation and test phases [41].

2.2. Long Short-Term Memory Network (LSTM)

The application of LSTM networks in calculating nonlinear functions and modeling data with a lower number of parameters has shown high performance [42]. These networks also use mathematical modeling and sophisticated methods to process data, thus finding hidden patterns in the data [42]. Figure 1 shows the structure of an LSTM unit. LSTM network is a method of deep learning and the main purpose of this method is to prevent zero gradients in the gradient descent method that occurs during the training of a shallow neural network model using the back propagation algorithm [42].

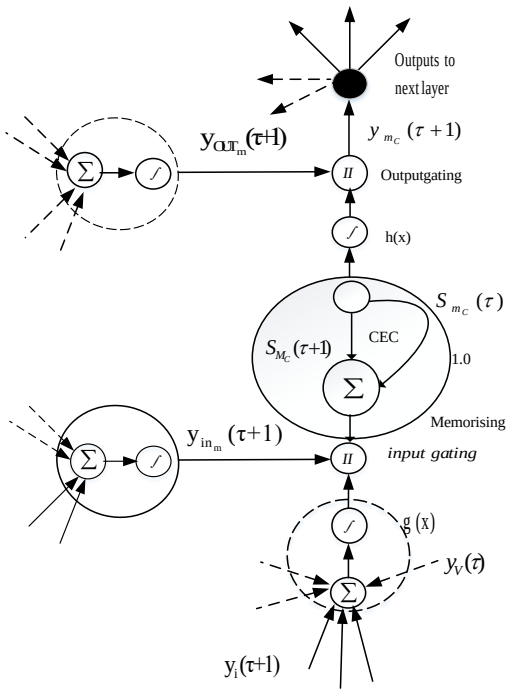


Figure 1. Structure of LSTM unit.

Each memory cell in the hidden layer of the LSTM network contains a self-connected recurrent edge. This edge has a weight of 1, which makes the gradient pass across many steps without exploding or vanishing [42]. LSTM consist of five main components: memory

block, memory cells, input gate, output gate, and forget gate. The whole computation can be defined by a series of equations as follows [42]:

$$i_t = \sigma\left(W^i H + b^i\right) \tag{2}$$

$$f_t = \sigma\left(W^f H + b^f\right) \tag{3}$$

$$o_t = \sigma(W^o H + b^o) \tag{4}$$

$$c_t = \tanh(W^c H + b^c) \tag{5}$$

$$m_t = f_t \odot m_{t-1} + i_t \odot c_t \tag{6}$$

$$h_t = \tanh(o_t \odot m_t) \tag{7}$$

σ is the sigmoid function, W^i, W^f, W^o, and W^c in $\mathbb{R}^{d\times 2d}$ are the recurrent weight matrices, and b^i, b^f, b^o, and b^c are the corresponding bias parameters. H is the concatenation of the new input x_t and the previously hidden vector h_{t-1} [42]:

$$H = \begin{bmatrix} Ix_t \\ h_{t-1} \end{bmatrix} \tag{8}$$

The key to LSTM is the cell state, i.e., memory vector m and m' which can remember long-term information. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. The gates including an input gate, a forget gate, an output gate, and a control gate is presented in Equations (2)–(7) by i, f, o, and c respectively. The input gate can decide how much input information enters the current cell. The forget gate can decide how much information be forgotten for the previous memory vector m_{i-1}, while the control gate can decide to write new information into the new memory vector m_i modulated by the input gate. The output gate can decide what information will be output from the current cell [42].

2.3. Adaptive Teaching–Learning-Based Optimization with Experience Learning

2.3.1. Teacher and Learner Phases in TLBO

Teaching-learning-based optimization was introduced in 2012 by [43] and it was proposed to solve large-scale non-linear problems. TLBO consists of two main phases, the teacher phase, and the learner phase [43]. If there are N learners in a classroom, then each learner is x_i can be supposed as a solution, and the teacher is the learner with the smallest objective function value. The teacher shares the knowledge with the learners, and each learner x_i learns knowledge from the teacher to improve his/her grade [43].

$$x_{i,new} = x_i + r(x_{teacher} - T_F x_{mean}) x_{i,new} = x_i + r(x_{teacher} - T_F x_{mean}) \tag{9}$$

$$x_{mean} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{10}$$

where $x_{i,new}$ is the ith learner after learning from the teacher, r is a decimal between 0 and 1, T_F is the teaching factor and its value is set 1 or 2 according to round [1 + rand (0, 1)], and x_{mean} is the average grade of the class. In the learner phase, each learner randomly selects another learner to interact with their knowledge and learns something new from another learner [43].

$$x_{i,new} = \begin{cases} (x_i + r)(x_i - x_j), & if\ x_i\ is\ better\ than\ x_j \\ (x_i + r)(x_j - x_i), & otherwise \end{cases} \tag{11}$$

2.3.2. Adaptive Selection

First, all learners are sorted according to their objective function values and divided into two subsets including fit solutions and inferior solutions [44]:

$$idx = sort(RMSE, 'ascend') \tag{12}$$

$$P = \begin{cases} fit\,solutions, & if\ idx \leq \frac{N}{2} \\ inferior\,solutions, & otherwise \end{cases} \tag{13}$$

Fit solutions will be selected for the teacher phase to improve convergence speed and exploitation capability, and inferior solutions will be selected for the learner phase to enhance the exploration capability.

2.3.3. Experience Learning

Two different learners x_{l1} and x_{l2} will be selected randomly in this phase of the ELATLBO algorithm while the current learner is x_i. The objective function value of x_{l1} and x_{l2} will be compared as follows [44]:

$$EL = x_{l_1} - x_{l_2} \tag{14}$$

The differential vector $(x_{l_1} - x_{l_2})$ is called experience learning (EL). When another learner uses the EL, the EL may guide the learner to a more promising search area [44]. A new teaching strategy proposed in this method is formulated as follows [44]:

$$x_{i,new} = x_i + r_1 \cdot (x_{teacher} - x_{IS}) + r_2 \cdot \left(x_{l_1} + x_{l_2}\right) \tag{15}$$

r_1 and r_2 are decimals between 0 and 1, and x_{IS} is an ordinary learner which is randomly selected from the inferior solutions.

A new learner strategy is proposed too which aims to enhance the population diversity and search more wide areas. Ordinary learners from the inferior solutions are selected in this phase and x_{mean} is utilized with the EL to improve the global search capability. The learner x_i updates using the following equation [44]:

$$x_{i,new} = x_i + r_1 \cdot (x_{mean} - x_i) + r_2 \cdot \left(x_{l_1} + x_{l_2}\right) \tag{16}$$

The pseudo-code of the ELATLBO algorithm is provided in Algorithm 1. The maximum number of function evaluations is Max_NFE, and NFE is the current number of function evaluations. The complexity of teacher or learner phase is $O(N)$ [44].

Algorithm 1: Pseudo-code of ELATLBO [44]

Input: Control parameters: Max_NFE, N
Set NFE = 0;
Initialize the population P randomly;
Evaluate the objective function values of the P;
NFE = NFE + N;
while NFE < Max_NFE do
Divide P into the better and ordinary learners using Equations (12) and (13);
for i = 1 to N do
if I < N/2 then
$x_{i,new} = x_i + r_1 \cdot (x_{teacher} - x_{IS}) + r_2 \cdot \left(x_{l_1} + x_{l_2}\right)$
else
$x_{i,new} = x_i + r_1 \cdot (x_{mean} - x_i) + r_2 \cdot \left(x_{l_1} + x_{l_2}\right)$
Re-evaluate the objective function values of updated;
Learners;
NFE = NFE + N
Output: optimal solution

2.4. The Hybrid Method for Demand Time Series Forecasting

In this study, the aim is to find optimal hyper-parameters of the LSTM model based on the ELATLBO method to forecast electricity demand. ELATLBO is the TLB method which enhanced with experiential learning and it can be utilized to obtain global solutions for continuous non-linear functions. Figure 2 illustrates the structure of the LSTM model in this paper and the fitness function is presented in Equation (17) subject to the limitations presented in Table 1.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_t} (y_i - \hat{y}_i)^2}{N_t}} \tag{17}$$

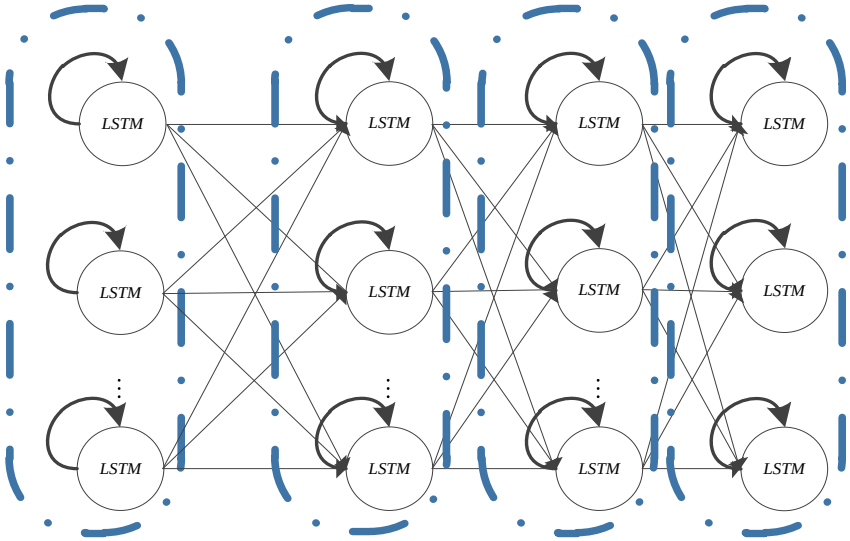


Figure 2. The proposed LSTM network structure.

Table 1. Defining hyperparameters of LSTM as design variables of the ELATLBO.

Design Variable	Hyperparameter	Limits
X_1	Hidden units number of 1st layer	$1 \leq X_1 \leq 200$
X_2	Hidden units number of 2nd layer	$1 \leq X_2 \leq 200$
X_3	Hidden units number of 3rd layer	$1 \leq X_3 \leq 100$
X_4	Hidden units number of 4th layer	$1 \leq X_4 \leq 100$
X_5	Maximum number of training epochs (LSTM)	$100 \leq X_6 \leq 500$
X_6	Learn rate	$0 < X_7 \leq 1$
X_7	Learn rate drop period	$1 \leq X_8 \leq 25$
X_8	Learn rate drop factor	$0 < X_9 \leq 1$

y_i is the vector of observed values and $\hat{y}_i$ is the vector of predicted values of electrical demand in the above equation. The goal of the proposed method in this paper is to minimize RMSE in the validation stage.

The flowchart of the proposed method for training and fine-tuning the LSTM model for demand forecasting showed in Figure 3.

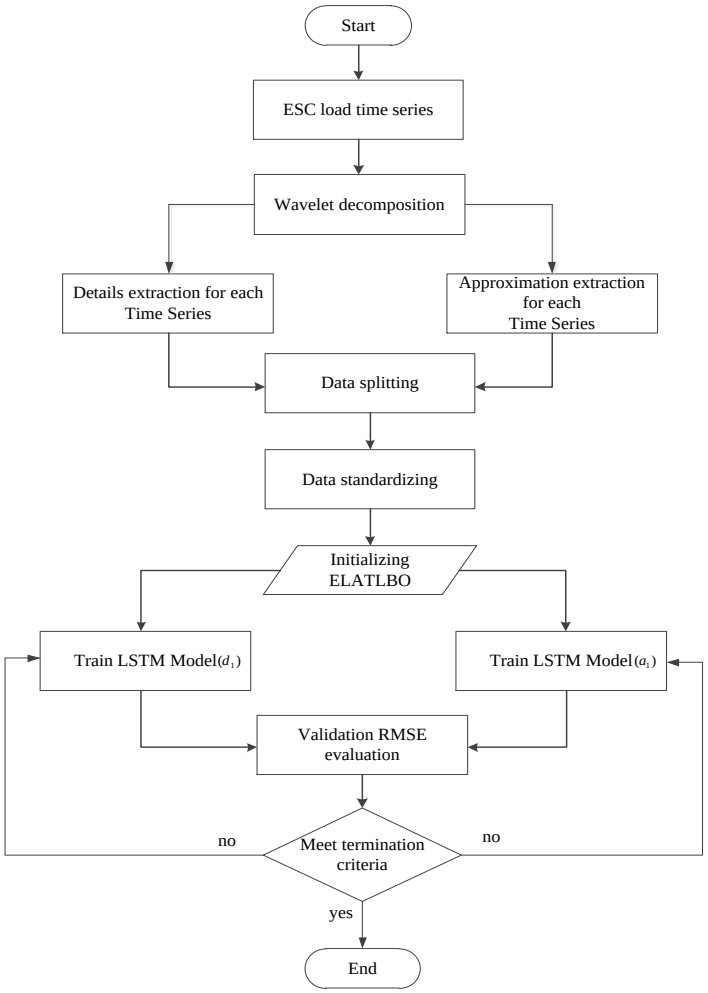


Figure 3. Flowchart of training and fine-tuning of the LSTM model for demand forecasting.

The demand time series of ESC is used in this study for demand time series forecasting of base metal industries micro-grids. Firstly, the original demand signal is decomposed into detail and approximation vectors using the Daubechies wavelet. The datasets consist of hourly recorded demand and the decomposed output signals have the same time points. In the next step, each vector will divide into three parts including training (50%), validation (19%), and test (31%) subsets.

Secondly, decomposed vectors d_1 and a_1 will be standardized and ELATLBO method parameters will be initialized. In the next step, LSTM models will be trained using initial values of hyper-parameters with a lag of one hour. When the training step is completed validation data is used to produce validation predictions and the fitness function i.e., RMSE will be evaluated in the current iteration. The algorithm will continue until one of the terminations criteria will be met.

Finally, the test data will be fed into trained and fine-tuned LSTM models to forecast new d_1 and a_1 components i.e., $\hat{d}_1$ and $\hat{a}_1$, and the final forecast vector will be produced by summing the elements of these two vectors.

2.5. Evaluation of Results

To assess the performance of a variety of forecasting methods, the mean absolute error (MAE), the mean squared error (MSE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE) are used as evaluation criteria. In addition, the

experiment is given from two aspects of single-step prediction and multi-step prediction, to verify its generalization performance [34].

$$\text{MAE} = \frac{1}{N_t}\sum_{i=1}^{N_t}|y_i - \hat{y}_i| \tag{18}$$

$$\text{MSE} = \frac{\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2}{N_t} \tag{19}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_t}(y_i - \hat{y}_i)^2}{N_t}} \tag{20}$$

$$\text{MAPE} = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{y - \hat{y}}{y}\right| \times 100 \tag{21}$$

3. Results and Discussion

This section of the paper is dedicated to presenting simulation results. The simulation was performed on a system with core i5-8250u CPU with 8 GB of RAM using MATLAB 2021 installed on window 10. ESC demand time series is selected to verify the performance of the coupled approach. The dataset used in this study includes hourly measured electrical demand of the ESC over 40 months from 21 March 2017, to 21 June 2020. The ESC demand signal is the only data that is used to predict next-hour demand in this study. Therefore, inputs in this study are 1-h lagged points of the decomposition of the original time series signal by the Daubechies wavelet i.e., details (d_1) and approximation (a_1) vectors. Figure 4 shows the original ESC demand signal and the details and approximation signals.

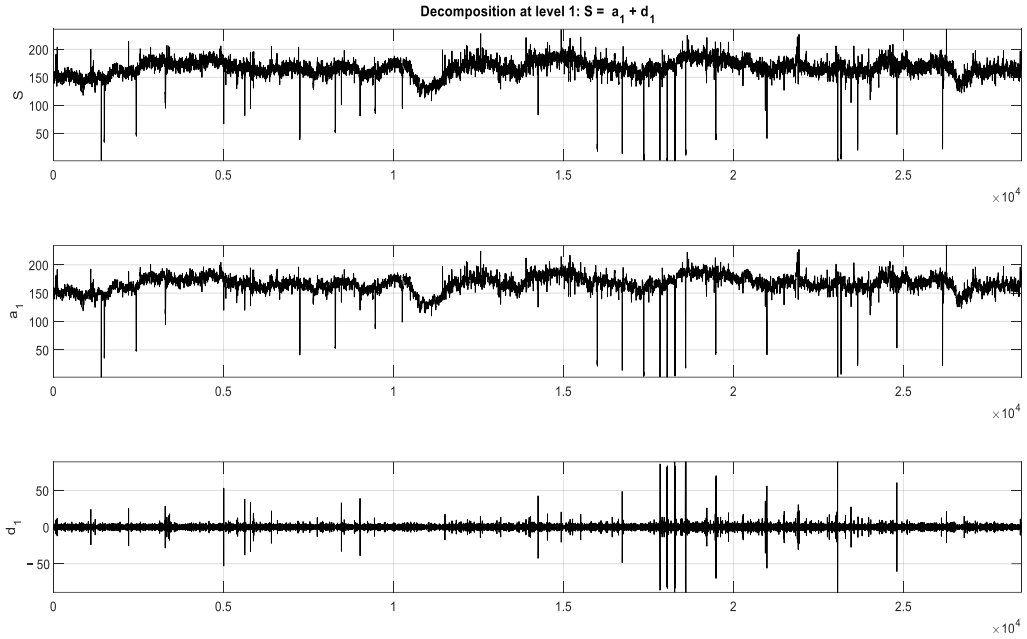


Figure 4. Illustration of the ESC demands time series, approximation, and details vectors obtained using the Daubechies wavelet.

In the next step of the proposed approach, the ELATLBO method will be used to tune the hyper-parameters of the LSTM model for training two separated forecasting models based on Algorithm 1 using the hyper-parameters of the LSTM defined as design variables of the ELATLBO presented in Table 1. One model will be trained to forecast the

d_1 signal and the other will be trained to forecast the a_1 signal. The control parameters Max_NFE and N in the ELATLBO algorithm are set to 10 and 20 respectively. The main advantage of using this method is the need to set fewer parameters in comparison with similar optimization algorithms.

The process of minimizing validation RMSE for d_1 and a_1 forecasting models is illustrated in Figures 5 and 6 respectively. The maximum number of runs of the ELATLBO is set at 150 and in each run, the LSTM training epochs are set to be between 100 to 500 in this study. Optimal hyper-parameters of LSTM obtained using the ELATLBO are presented in Table 2 and the prediction results are illustrated in Figure 7.

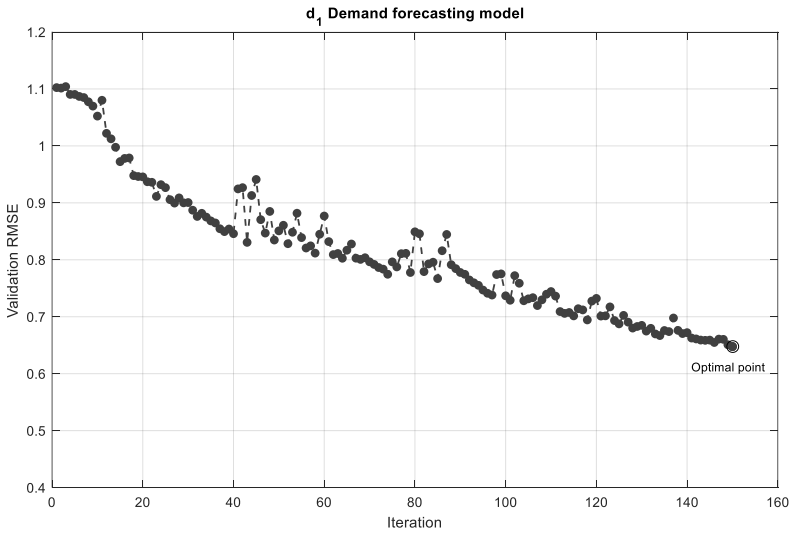


Figure 5. Minimizing validation RMSE of the d_1 forecasting model.

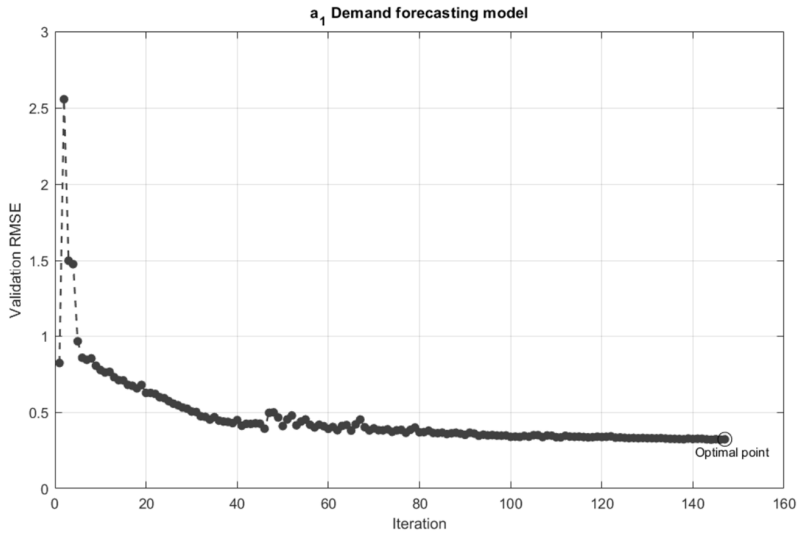


Figure 6. Minimizing validation RMSE of the a_1 forecasting model.

After standardization, the test subsets of d_1 and a_1 vectors are inputted into the trained models to produce initial forecast signals. The predicted signal illustrated in Figure 7 is the result of the summation of d_1 and a_1 vectors. To evaluate the obtained results, the proposed method has been compared with other methods including Support Vector Machine (SVM), Decision Tree, Boosted Tree, and Random Forest while the hyper-parameters of the aforementioned methods are optimized using the Bayesian Optimization (BO) method. The obtained results are evaluated in Table 3 and Figures 8–11. Comparing the results of Table 3 shows that the proposed method performs well while the parameters set in the proposed method are much fewer than similar methods.

Table 2. Optimal hyperparameters of LSTM for demand forecasting obtained by the ELATLBO.

Design Variable	Hyperparameter	d_1	a_1
X_1	Hidden units number of 1st layer	195	200
X_2	Hidden units number of 2nd layer	181	189
X_3	Hidden units number of 3rd layer	92	90
X_4	Hidden units number of 4th layer	77	82
X_5	Maximum number of training epochs	500	500
X_6	Learn rate	0.00489	0.00501
X_7	Learn rate drop period	22	25
X_8	Learn rate drop factor	0.2010	0.1869

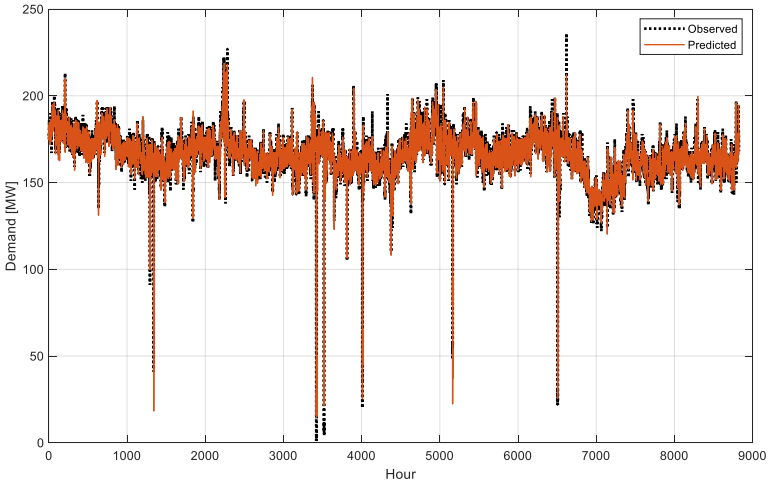


Figure 7. Forecasted and predicted demand of the ESC using the proposed method.

Table 3. Test results.

Method	Optimization Method	MAE	MAPE	MSE	RMSE	Standard Deviation
W-LSTM	ELATLBO	3.023	1.8161	39.515	6.2861	12.9090
Gaussian SVM	BO	3.2191	1.9339	53.895	7.3413	13.4734
Decision Tree	BO	3.1224	1.8758	49.847	7.0603	13.5507
Boosted Tree	BO	3.2901	1.9765	48.299	6.9498	13.8292
Random Forest	BO	3.0611	1.839	48.513	6.9651	13.6558

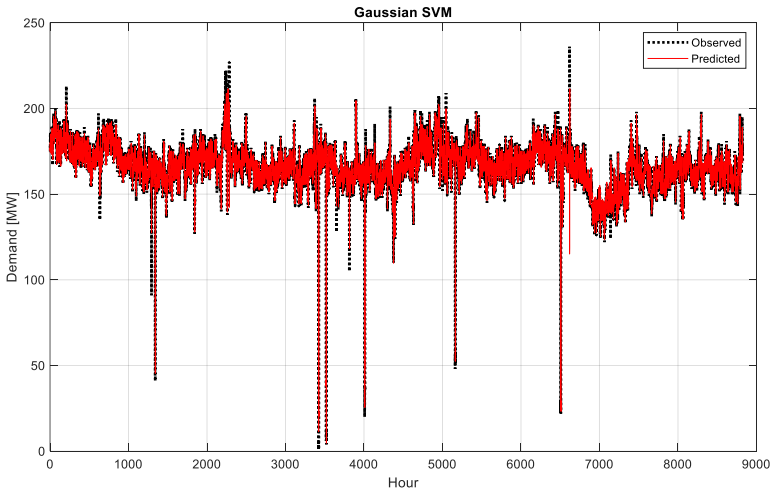


Figure 8. The ESC demand Forecasting using the Gaussian SVM model optimized using BO.

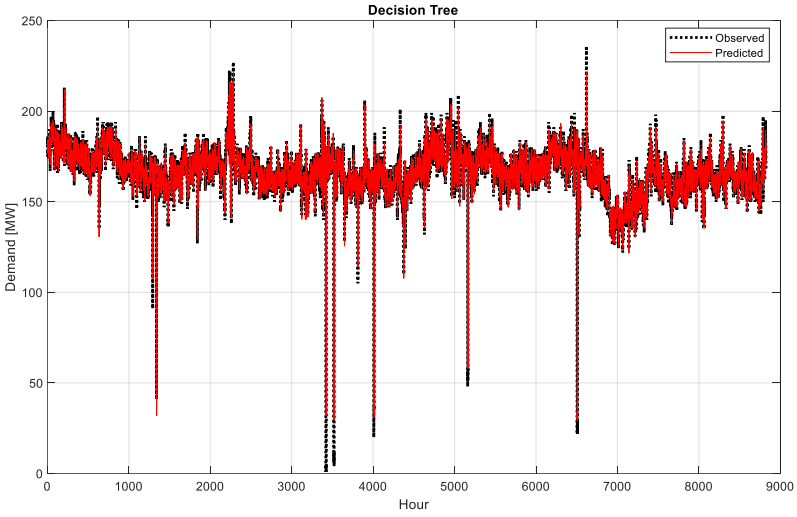


Figure 9. The ESC demand Forecasting using the Decision Tree model optimized using BO.

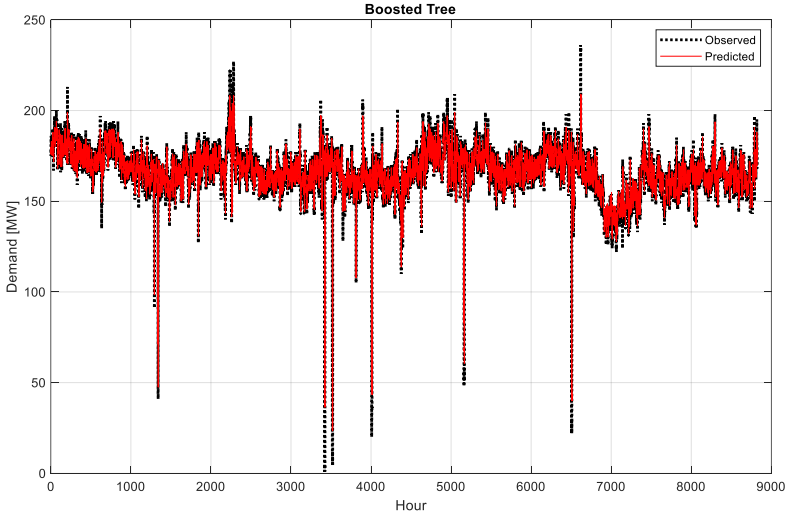


Figure 10. The ESC demand Forecasting using the Boosted Tree model optimized using BO.

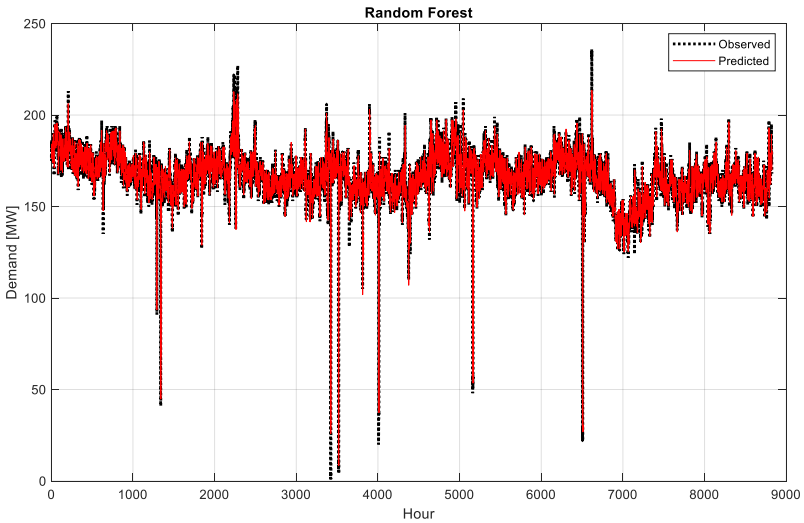


Figure 11. The ESC demand Forecasting using the Random Forest model optimized using BO.

The illustrated results in Table 3 and Figures 7–11 show that the proposed method based on wavelet analysis and optimization of hyper-parameters of LSTM based on the ELTABO method is an effective method for training of accurate demand forecasting models. The error index values obtained and the presented plots show the high accuracy of the demand forecasting. The obtained MAPE value shows that mean absolute percentage error is less than 2% in the entire forecast period, which is considered a favorable result.

4. Conclusions

Demand forecasting is of great importance in balancing the demand-supply chain. The basic metal industry is one of the energy-consuming industries in Iran and specifically in Isfahan Province. Accurate electrical demand forecasting of basic metal industries to plan electrical generation is the first step in balancing the electrical supply chain. A hybrid method is proposed in this paper to forecast the demand time series of the metal industries' micro-grid. The coupled approach consists of wavelet decomposition in the first step. Then the output vectors of the wavelet decomposition are split into three parts. The training subsets and the validation subsets are used in the training and fine-tuning of the LSTM model using the ELATLBO method. The ESC dataset used in this study for electrical demand forecasting includes 24-h daily over 40 months from 21 March 2017, to 21 June 2020. For evaluating the obtained results, a comparison has been made with other methods including SVM, Decision Tree, Boosted Tree, and Random Forest which are optimized using the Bayesian Optimization (BO) method. The obtained results with minimum available data (only demand time series) show that the performance of the coupled approach is well while the parameters set in the proposed method are much fewer than similar approaches. In the continuation of the research carried out in this study, it is possible to consider and study an optimization algorithm with high adaptability and minimizing tunable parameters exclusively for the optimization of regression methods.

Author Contributions: Conceptualization, All authors; methodology, S.M., R.M.A., G.S. and M.M..; validation, S.M. and M.M.; investigation, All authors; resources, S.M. and M.M.; data curation; All authors, writing—original draft preparation, S.M. and M.M.; writing—review and editing, All authors; supervision, R.M.A., G.S., M.M. and S.M.K.; validation, S.M. and M.M.; project administration, R.M.A., G.S., M.M. and S.M.K.; funding acquisition, This research received no external funding. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

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

References

1. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
2. Domhan, T.; Springenberg, J.T.; Hutter, F. Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
3. Mastrocinque, E.; Ramírez, F.J.; Honrubia-Escribano, A.; Pham, D.T. An AHP-based multi-criteria model for sustainable supply chain development in the renewable energy sector. *Expert Syst. Appl.* **2020**, *150*, 113321. [CrossRef]
4. Sarah, J.; Khalili-Damghani, K. Fuzzy type-II De-Novo programming for resource allocation and target setting in network data envelopment analysis: A natural gas supply chain. *Expert Syst. Appl.* **2019**, *117*, 312–329.
5. Abolghasemi, M.; Beh, E.; Tarr, G.; Gerlach, R. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Comput. Ind. Eng.* **2020**, *142*, 106380. [CrossRef]
6. Abedinia, O.; Amjady, N.; Zareipour, H. A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems. *IEEE Trans. Power Syst.* **2016**, *32*, 62–74. [CrossRef]
7. Lin, W.-M.; Tu, C.-S.; Yang, R.-F.; Tsai, M.-T. Particle swarm optimisation aided least-square support vector machine for load forecast with spikes. *IET Gener. Transm. Distrib.* **2016**, *10*, 1145–1153. [CrossRef]

8. Jiang, H.; Zhang, Y.; Muljadi, E.; Zhang, J.J.; Gao, D.W. A Short-Term and High-Resolution Distribution System Load Forecasting Approach Using Support Vector Regression with Hybrid Parameters Optimization. *IEEE Trans. Smart Grid* **2016**, *9*, 3341–3350. [CrossRef]
9. Liu, Y.; Sun, Y.; Infield, D.; Zhao, Y.; Han, S.; Yan, J. A Hybrid Forecasting Method for Wind Power Ramp Based on Orthogonal Test and Support Vector Machine (OT-SVM). *IEEE Trans. Sustain. Energy* **2016**, *8*, 451–457. [CrossRef]
10. Nowotarski, J.; Weron, R. Computing electricity spot price prediction intervals using quantile regression and forecast averaging. *Comput. Stat.* **2015**, *30*, 791–803. [CrossRef]
11. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. *IEEE Trans. Smart Grid* **2015**, *8*, 730–737. [CrossRef]
12. Zhang, W.; Quan, H.; Srinivasan, D. An improved quantile regression neural network for probabilistic load forecasting. *IEEE Trans. Smart Grid* **2018**, *10*, 4425–4434. [CrossRef]
13. Liu, Q.; Shen, Y.; Wu, L.; Li, J.; Zhuang, L.; Wang, S. A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting. *CSEE J. Power Energy Syst.* **2018**, *4*, 226–237. [CrossRef]
14. Zeng, P.; Jin, M. Peak load forecasting based on multi-source data and day-to-day topological network. *IET Gener. Transm. Distrib.* **2018**, *12*, 1374–1381. [CrossRef]
15. Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Trans. Smart Grid* **2017**, *9*, 5271–5280. [CrossRef]
16. 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 Trans. Smart Grid* **2017**, *10*, 841–851. [CrossRef]
17. Xu, F.Y.; Cun, X.; Yan, M.; Yuan, H.; Wang, Y.; Lai, L.L. Power Market Load Forecasting on Neural Network with Beneficial Correlated Regularization. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5050–5059. [CrossRef]
18. Rafiei, M.; Niknam, T.; Aghaei, J.; Shafie-Khah, M.; Catalão, J.P. Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine. *IEEE Trans. Smart Grid* **2018**, *9*, 6961–6971. [CrossRef]
19. Yu, Z.; Niu, Z.; Tang, W.; Wu, Q. Deep Learning for Daily Peak Load Forecasting–A Novel Gated Recurrent Neural Network Combining Dynamic Time Warping. *IEEE Access* **2019**, *7*, 17184–17194. [CrossRef]
20. Ye, C.; Ding, Y.; Wang, P.; Lin, Z. A data-driven bottom-up approach for spatial and temporal electric load forecasting. *IEEE Trans. Power Syst.* **2019**, *34*, 1966–1979. [CrossRef]
21. Ouyang, T.; He, Y.; Li, H.; Sun, Z.; Baek, S. Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *3*, 127–136. [CrossRef]
22. Chen, K.; Chen, K.; Wang, Q.; He, Z.; Hu, J.; He, J. Short-Term Load Forecasting with Deep Residual Networks. *IEEE Trans. Smart Grid* **2018**, *10*, 3943–3952. [CrossRef]
23. Deng, Z.; Wang, B.; Xu, Y.; Xu, T.; Liu, C.; Zhu, Z. Multi-Scale Convolutional Neural Network with Time-Cognition for Multi-Step Short-Term Load Forecasting. *IEEE Access* **2019**, *7*, 88058–88071. [CrossRef]
24. Wang, Y.; Zhang, N.; Chen, Q.; Kirschen, D.S.; Li, P.; Xia, Q. Data-Driven Probabilistic Net Load Forecasting with High Penetration of Behind-the-Meter PV. *IEEE Trans. Power Syst.* **2017**, *33*, 3255–3264. [CrossRef]
25. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*; 2011; p. 24.
26. Kim, T.-Y.; Cho, S.-B. Particle swarm optimization-based CNN-LSTM networks for forecasting energy consumption. In Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC), Wellington, New Zealand, 10–13 June 2019.
27. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
28. Suganuma, M.; Shirakawa, S.; Nagao, T. A genetic programming approach to designing convolutional neural network architectures. In Proceedings of the genetic and evolutionary computation conference, Berlin, Germany, 15–19 July 2017.
29. Lv, L.; Kong, W.; Qi, J.; Zhang, J. An improved long short-term memory neural network for stock forecast. *MATEC Web Conf.* **2018**, *232*, 01024. [CrossRef]
30. Rokhsatyazdi, E.; Rahnamayan, S.; Amirinia, H.; Ahmed, S. Optimizing LSTM Based Network for Forecasting Stock Market. In Proceedings of the 2020 IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 19–24 July 2020.
31. Khan, M.A.; Saqib, S.; Alyas, T.; Rehman, A.U.; Saeed, Y.; Zeb, A.; Zareei, M.; Mohamed, E.M. Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access* **2020**, *8*, 116013–116023. [CrossRef]
32. Wu, D.; Xu, Y.T.; Jenkin, M.; Wang, J.; Li, H.; Liu, X.; Dudek, G. Short-term Load Forecasting with Deep Boosting Transfer Regression. In Proceedings of the ICC 2022-IEEE International Conference on Communications, Seoul, Korea, 16–20 May 2022.
33. Han, X.; Su, J.; Hong, Y.; Gong, P.; Zhu, D. Mid-to Long-Term Electric Load Forecasting Based on the EMD–Isomap–Adaboost Model. *Sustainability* **2022**, *14*, 7608. [CrossRef]
34. Alsharekh, M.F.; Habib, S.; Dewi, D.A.; Albattah, W.; Islam, M.; Albahli, S. Improving the Efficiency of Multistep Short-Term Electricity Load Forecasting via R-CNN with ML-LSTM. *Sensors* **2022**, *22*, 6913. [CrossRef]
35. 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]
36. Xu, A.; Tian, M.-W.; Firouzi, B.; Alattas, K.A.; Mohammadzadeh, A.; Ghaderpour, E. A New Deep Learning Restricted Boltzmann Machine for Energy Consumption Forecasting. *Sustainability* **2022**, *14*, 10081. [CrossRef]

37. Alotaibi, M.A. Machine Learning Approach for Short-Term Load Forecasting Using Deep Neural Network. *Energies* **2022**, *15*, 6261. [CrossRef]
38. Kondaiah, V.Y.; Saravanan, B. Short-Term Load Forecasting with a Novel Wavelet-Based Ensemble Method. *Energies* **2022**, *15*, 5299. [CrossRef]
39. Duan, Y. A Novel Interval Energy-Forecasting Method for Sustainable Building Management Based on Deep Learning. *Sustainability* **2022**, *14*, 8584. [CrossRef]
40. Bai, W.; Zhu, J.; Zhao, J.; Cai, W.; Li, K. An Unsupervised Multi-Dimensional Representation Learning Model for Short-Term Electrical Load Forecasting. *Symmetry* **2022**, *14*, 1999. [CrossRef]
41. Strang, G.; Nguyen, T. *Wavelets and Filter Banks*; Wellesley-Cambridge: Cambridge, MA, USA, 1996.
42. Liu, Y.; Qin, Y.; Guo, J.; Cai, C.; Wang, Y.; Jia, L. Short-term forecasting of rail transit passenger flow based on long short-term memory neural network. In Proceedings of the 2018 International Conference on Intelligent Rail Transportation (ICIRT), Singapore, 12–14 December 2018.
43. Rao, R.V.; Savsani, V.J.; Vakharia, D. Teaching–learning-based optimization: An optimization method for continuous non-linear large scale problems. *Inf. Sci.* **2012**, *183*, 1–15. [CrossRef]
44. Mi, X.; Liao, Z.; Li, S.; Gu, Q. Adaptive teaching–learning-based optimization with experience learning to identify photovoltaic cell parameters. *Energy Rep.* **2021**, *7*, 4114–4125. [CrossRef]