Tanzania Journal of Science 49(1): 130-139, 2023
ISSN 0856-1761, e-ISSN 2507-7961

Short-Term Energy Load Forecasting Model with Sample Filtering

Ruthbetha Kateule*, Lucas A. Sakala and Mahadia Tunga
Department of Computer Science and Engineering, University of Dar es Salaam,
P.O. Box 33335, Dar es Salaam, Tanzania.
Email addresses: rkateule@udsm.ac.tz, lucasalicksakala@gmail.com,
mahadiatunga@udsm.ac.tz
**Corresponding author*
Received 30 Nov 2022, Revised 20 Feb 2023, Accepted 21 Feb 2023 Published Mar 2023
DOI: https://dx.doi.org/10.4314/tjs.v49i1.12

Abstract
Short-term energy load forecasting is a crucial task in the power smart grid, which enables the power utilities to understand the future energy demands and plans to attain the demand and supply equilibrium, thereby optimizing power deployment and reducing power losses. Several techniques have been implemented to enhance energy load forecasting. However, the nonlinear nature of the data collected in the smart grid makes it difficult to attain 100% energy load forecasting accuracy. For instance, the Deep Feedforward Neural Networks model based on Input Attention Mechanism and Hidden Connection Mechanism has a mean absolute percentage error of 3.17%; model based on Sequence to Sequence Recurrent Neural Network with Attention had a mean absolute percentage error of 2.7%. The model based on Deep Recurrent Neural Networks with Levenberg–Marquardt backpropagation algorithm had a mean absolute percentage error of 0.58; and Deep Feedforward Neural Network with sample weights model had 3.22 % as root mean squared error. To improve energy load forecasting accuracy, this work proposed a model based on Deep Recurrent Neural Networks and sample filtering, which provides an exhaustive elucidation for modelling a sophisticated stochastic relationship between the input and output features. Deep Recurrent Neural Networks have proven to be good at modelling the nonlinearities in data of different fields and are mostly used in energy load forecasting to reduce forecasting error and a high degree of overfitting. Sample filtering is achieved through the use of K-Means clustering which determines the number of clusters to be used in the model. Findings from the study showed that by employing Deep Recurrent Neural Networks and sample filtering, the short-term energy load forecasting accuracy is improved in reference to mean absolute percentage error and root mean squared error of 0.31% and 1.014, respectively. As a result of the reduction in error, the energy demand and supply chain equilibrium are enhanced, thereby optimizing power deployment and reducing power losses.

Keywords: Machine learning, Neural networks, Sample filtering, Smart grid, Short-term energy forecasting.

Introduction
The availability of data and growing technology have aided in transforming the traditional electrical grid into a smart grid. The smart grid is composed of several elements, i.e., smart meters, measurement units, and sensors. These elements produce an enormous amount of data at a high velocity to support the smart grid. A power smart grid is an electrical grid that incorporates a diversity of operations, energy measures, and energy resources (Figure 1). It comprises energy generation, transmission and distribution, and consumption phases.

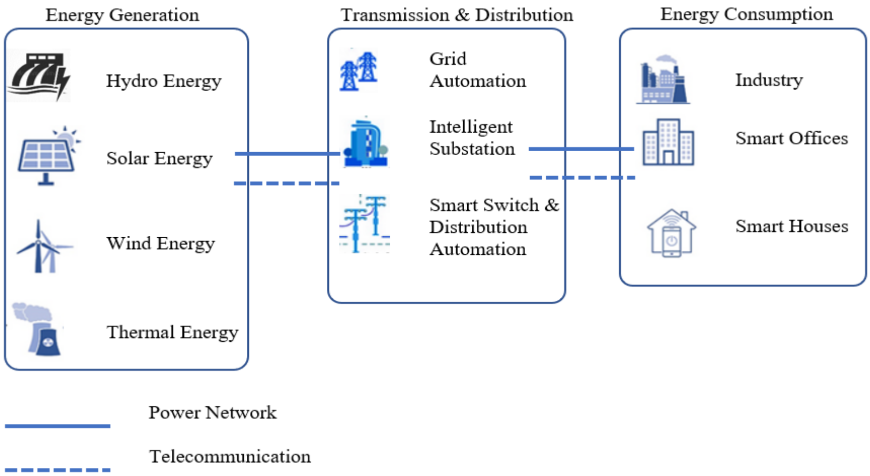

Figure 1: Energy smart grid overview (Mujeeb et al. 2018).

Smart grid (SG) enhances the productivity and dependability of energy production, distribution, and utilisation (Mujeeb et al. 2018). The power smart grid instigates a central role in the state-of-the-art energy infrastructure. The smart meters measure electricity utilisation at every minor interval and impart energy suppliers, ensuring the generation of an enormous quantity of data. Attributable to the hereness of these data, many innovative schemes are enforced, which include real-time pricing.

By utilising data analytics tools, hidden patterns in the generated data can be revealed, which show the correlations between different features (Mujeeb et al. 2018). Data analytics tools leverage neural network algorithms to extract prized information from the data (Marinakis 2020). The unveiled information can be used in the management of the generation and consumption of energy through Load Forecasting (LF). LF is a technique to predict future energy requirements to attain demand and supply equilibrium (Wei et al. 2019, Syed et al. 2020). LF plays a pivotal part in energy management as it facilitates generation, distribution, and energy consumption (Zhang et al. 2018). Both meteorological data and day type play an essential role in LF as they determine energy consumption (Cai et al. 2020). For instance, a day type can be used to determine the consumption of energy based on non-working and working days (Cai et al. 2020). LF is an integral decision-supporting tool that underpins proficient energy management in the smart grid (Syed et al. 2020). Accurate LF ensures the poise between generation and demand, refraining from the waste of resources and remodelling the stability of power systems.

Power utility companies need accurate and reliable power LF to support the stability and safety of the power supply. The least progression in the accuracy of LF for the energy realm expedites big gleanings with considerable environmental and economic gains. Accurate LF is therefore required to enable the utilities to understand the future energy arduous and plan their infrastructures to meet the demands (Syed et al. 2020). However, the electricity load is nonlinear with a high level of volatility which makes it difficult to attain 100% energy load forecasting accuracy. Since the nonlinearity makes the power supply infrastructures to be very sensitive to severe weather-induced load variations. Nevertheless, the erratic nature of energy consumption give-rise-to compelling skepticism and decoherence to the power smart grid (Mohammad and Kim 2020). This may bring about an extensive strain on the power smart grid and has huge impacts on energy generation and distribution. Thus, accurate LF becomes a critical component of smart grid for effective management and operations of power supply infrastructures in the microgrid.

Related work

According to Narayan and Hipel (2017) and Groß et al. (2021), entrenched on the time scale and their purposes, LF models are indexed into three clusters:

(i) Short-term load forecast (STLF): Span of STLF ranges from some minutes to a week. STLF targets economic alacrity and excellent generation engagement, guaranteeing security assessment and real-time control.

(ii) Medium-term load forecast (MTLF): Span of MTLF is from a week to a year. MTLF targets maintenance arrangements, designation of load dispatch, and price resolution to attain demand and supply equilibrium.

(iii) Long-term load forecast (LTLF): Span of LTLF varies from a year and beyond. LTLF targets outlining system augmentation, which embraces electricity generation, transmission, and distribution.

This work focuses on the STLF approaches as they are extremely important for the real-time affairs of power systems to avert the far-reaching consequences of power fizzles. STLF is imperative for the productive handling of power systems and acts as the foundation for formulating start-ups and shutdown schedules, which present a crucial part of the power system's automatic control (Zhang et al. 2020). The prediction process of STLF approaches are influenced by various factors including; (i) Time is a critical factor for STLF which includes time of day, the day of the week, holidays, and weekends, (ii)Weather factors, i.e., temperature, humidity, precipitation, etc., (iii) Size of the house, and (iv) Global factors such as diseases, etc. (Behnam et al. 2021).

Besides, most of the existing LF techniques focused on time series and weather factors which are not limited to time series (Papadopoulos and Karakatsanis 2015), regression methods, and neural networks (Pramono et al. 2019), expert systems (Qiu et al. 2017), and support vector machines (Raza and Khosravi 2015). Hui et al. (2017) proposed a deep neural network model which confined several hidden layers that extracted deep characteristics of data. In this model, a genetic algorithm was utilized in optimizing the weights and thresholds on the Deep Neural Networks (DNN). However, the learning algorithm of such a model is very slow and computationally expensive especially with nonlinear data and large datasets. The training rate could be accelerated through the use of advanced learning algorithms such as Deep Recurrent Neural Network (DRNN) which could eventually reduce the computational cost (Zhang et al. 2019).

Zhang et al. (2020) proposed a Recurrent Neural Network (RNN)-based LF model to improve accuracy. The model employed the Input Attention Mechanism (IAM) and Hidden Connection Mechanism (HCM) with 287.51 as Root Mean Squared Error (RMSE) Megawatt, 3.17 Mean Absolute Percentage Error (MAPE) (%), and 23.18 as convergent time(s). HCM improved converging speed through the use of residual connections which further helped to improve the model's efficiency. However, the performance of the model was low since training samples were not selected based on the similarity with the forecast day, sample filtering could have further improved the training results.

Xu et al. (2020) proposed a STLF model established on an ensemble residual network with MAPE (%) [2.705, 4.116, 5.599, 6.774] with prediction length (time steps) of N = [12, 48, 120, 288], respectively. The proposed model was built on a two-stage network structure. The first network stage was made up of entirely connected layers of which its exploratory results were fed to the second stage. In the second stage, a modified residual network did the final predictions. In this model, learning rate decay was utilized to enhance the accuracy of the proposed model. However, the model did not perform well in a longer input sequence since only electricity load and temperature were considered as input features of the model. More factors should be considered to improve the accuracy of the model.

Even though various ML models have been introduced for STLF in recent years, no model has attained 100% prediction accuracy; thus, the LF models still need to be improved. Moreover, this work proposes a STLF model with sample filtering to improve load forecasting accuracy through the use of advanced learning algorithms mainly Deep Recurrent Neural Network (DRNN) and sample filtering. The DRNN has a better learning rate especially in handling nonlinear data due its capacity in estimating anonymous dynamics of nonlinear systems compared to traditional DNN (Han et al. 2015). Sample filtering enhances training ability by removing the power load stochastic noise in power load data and eventually ensuring samples having a significant degree of affinity to the prediction day feature are used in training (Huang et al. 2017).

Materials and Methods

Experimentation

The research adopts an experimental design. The core focus of the experimentation was to obtain the MAPE of the proposed model, referred to as Model 1, and compare it with the other models in the study, which are: the Deep Feedforward Neural Network (DFNN) model based on the Input Attention Mechanism (IAM) and Hidden Connection Mechanism (HCM), referred to as Model 2; model based on Sequence to Sequence Recurrent Neural Network (S2S RNN) with Attention referred to as Model 3; DFNN with sample weights model referred to as Model 4 and DRNN with Levenberg–Marquardt (LM) backpropagation algorithm designated as Model 5. The study's first phase involved acquiring and pre-processing raw data (Figure 2). It further involved the identification of features to be used in the model. The second phase was aimed at designing and implementing the proposed model. Lastly, the proposed model was evaluated through MAPE and RMSE.

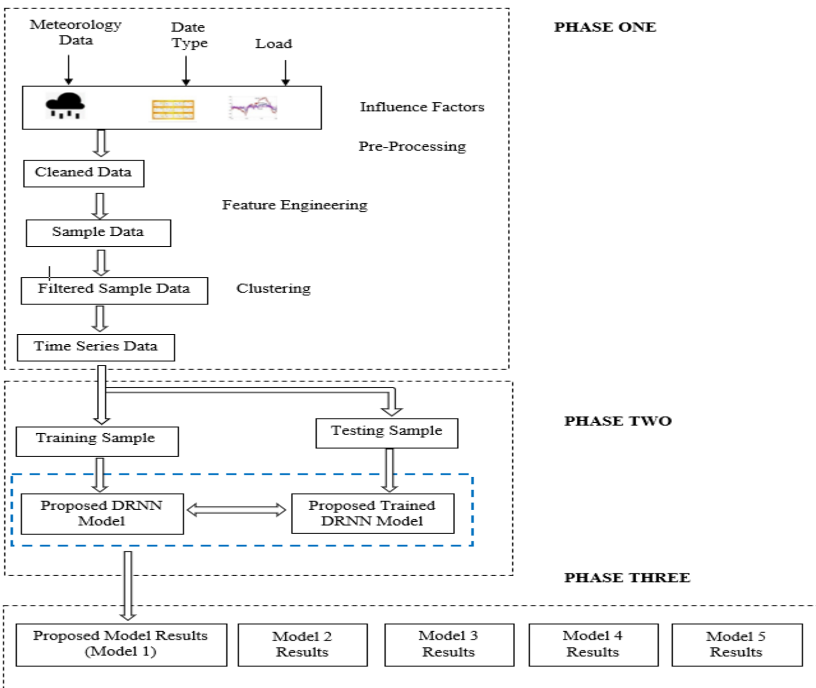

Figure 2: Research design.

Data acquisition and pre-processing

The work utilised the London smart meter dataset, as it is readily available for research objectives on Kaggle.com. It is a restructured edition of the London Energy dataset, which accommodates 5,567 London households' samples of the energy consumption readings that participated in the UK Power Networks led Low Carbon London project between November 2011 and February 2014. Considering that the smart meters were progressively rolled out in this project, the number of households whereupon the data were recorded differed for different days. To avert erroneous analysis, this work has considered the average day-level energy usage per household to standardise data. Also, considering the weather data as an integral factor for energy consumption, the daily weather data was acquired from the DarkSky API. UK holiday dataset was also utilised, considering the relevance of holiday features in energy consumption. Exploratory Data Analysis (EDA) and feature engineering were performed on weather data to identify the relevant features for developing the STLF model based on DRNN and sample filtering.

During the pre-processing stage, data cleaning, followed by imputation techniques to curate the dataset and fill in missing values. Standardization was applied to the datasets to deal with numerical dissimilarities for feature variables such as wind speed, temperature, and load. Standardisation was achieved by transforming the feature values to have zero mean and unit-variance through equation (1).

$$x\tilde{} = \frac{x-\mu}{\sigma} \quad (1)$$

Where: x denotes the authentic feature vector, μ represents the mean of the respective feature vector, σ is the standard deviation for the feature vector, and $x\tilde{}$ denotes the normalised feature vector. As the training and forecast day samples are standardized, denormalization was performed after consumption data output through the denormalisation equation shown in equation (2).

$$x = x^{*}\sigma + \mu \cdot \quad (2)$$

Model fitting

The 70:30 training/test split ratio was used. The mode fitting of the DRNN involved repeatedly adjusting the weights and biases to minimise the value of the loss function (Figure 3). The number of input variables determines the number of units in the input layer of the model whose output regulates the number of output layer units. The learning algorithm is pivotal in establishing the weight values to yield the most favourable prediction results. Each unit in the network is defined by its weight, bias, and activation function, which performs the nonlinear transformations. Once the hidden layers complete the processing, the final output is sent to the output layer. In a neural network, a technique known as backpropagation is utilised to reduce the loss function value. This is achieved by updating the weights and biases of the units based on the prediction error. Once the training process is complete, the best-fit weight and biases are utilised in the predictions.

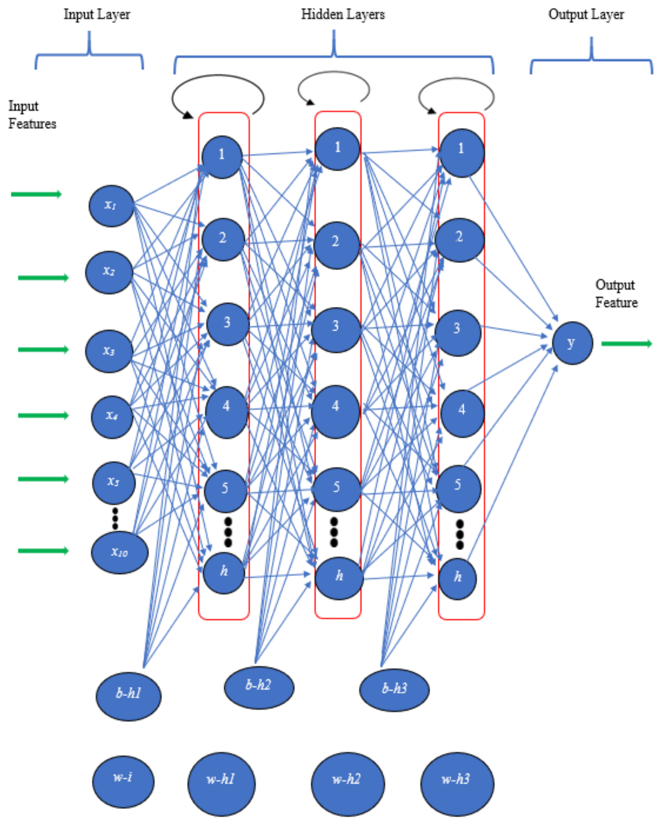

Figure 3: Deep recurrent neural network.

Model optimisation

As pointed out, the neural network was trained by iteratively adjusting the weights and biases to reduce the loss function through the learning algorithm. The loss function adopted in this work is the mean squared error presented in equation (3). Furthermore, techniques such as early stopping and dropout were implemented to reduce the overfitting of the model, thereby increasing its efficiency.

$$L_0 = \frac{1}{2m} \sum_m \| \hat{y} - y \|^2 \qquad (3)$$

Where: m denotes the number of samples, $\hat{y}$ represents the DRNN output value, and y denotes the actual sample load. A regularisation strategy is introduced to avert the DRNN overfitting, and the loss function L is then transformed, as shown in equation (4).

$$L = \frac{1}{2m} \sum_m \left(\| \hat{y} - y \|^2 + \lambda \| w \|^2 \right) \qquad (4)$$

Where: λ denotes the regularisation coefficient, and w represents the weight coefficient.

The training process utilised the learning rate decay and the Root Mean Square Propagation (RMSProp). The learning rate decay technique prevented the algorithm from oscillating back and forth, thereby reducing the learning rate. The RMSProp enabled the adaptive updating of the parameters by keeping the moving average of the square gradient constant. An activation function of the neural network serves as a decisive guideline to transfer weighted inputs to engender the network outputs. The stochastic activation function enables neural networks to implement sophisticated computations. Rectified Linear Activation Function (ReLU) shown in equation (5) is the extensively utilised activation function in DNN, which was used in this work. ReLU is a linear function that returns a value of zero if the input is negative and returns the input value for each positive input. In doing so, the vanishing gradient problem is averted.

$$f(x) = \max(0, x) \quad (5)$$

Where: x denotes the neuron input.

Model evaluation

This study utilised experiments for an empirical evaluation of the performance of the proposed model. The core focus of the experimentation was to obtain the MAPE metric and RMSE of the proposed model, referred to as Model 1, and compare it with the four other models in the study, which are: the Deep Feedforward Neural Network (DFNN) model based on the Input Attention Mechanism (IAM) and Hidden Connection Mechanism (HCM), referred to as Model 2; model based on Sequence to Sequence Recurrent Neural Network (S2S RNN) with Attention referred to as Model 3; DFNN with sample weights model referred to as Model 4 and DRNN with Levenberg–Marquardt (LM) backpropagation algorithm designated as Model 5. MAPE is the average of the absolute percentage errors of forecasts.

The contrast between the true and forecasted values is regarded as a forecasting error. RMSE presents the error in terms of percentages enabling a more straightforward interpretation. The complication of positive and negative error cancellation is eliminated as only absolute percentage errors are considered. Notably, forecasting accuracy increases with the reduction in MAPE values. MAPE of the proposed model was evaluated against Model 2, Model 3, and Model 5. RMSE is the square root of Mean Square Error and was employed in evaluating the recommended model to Model 4.

Results and Discussion

Selected features for the proposed short term load forecasting model

Features were selected based on their impacts in the consumption of electric power. Table 1 shows the correlations between weather variables and energy consumption. Pressure and moon phase have minimal correlations with energy consumption, hence removed from the final weather dataset. Weather factors such as temperature, humidity, pressure, and wind speed are typically among the most significant features in setting the STLF. A summer heatwave will spur consumers to run their air conditioners more and drive up the power demands. Similarly, a period of extreme cold in the winter will prompt consumers' heating equipment to run more frequently. This includes equipment such as electric heat pumps and blowers that circulate warm air throughout homes and businesses. Moderate weather in the spring and fall, on the other hand, tends to minimize the uses of such equipment and reduces the power demands.

Considering the effects of non-working-day or working-day on energy consumption, the holiday indicator was also added to the final dataset. On weekdays, electricity consumption is high while on Saturday and Sunday electricity consumption is low, and similarly on other social holidays. Based on these observations, the "Holiday Indicator" was added to draw this effect. Since the accuracy of the model is correlated to the affinity between the training samples and the prediction day (Cai et al. 2020), a new feature called energy cluster was also created from the average energy consumption. The new feature was used in the filtering of the sample data based on their similarity by creating clusters using K-Means clustering. Finally, Table 2 summarises the selected features for the proposed model; meteorological data which included the maximum temperature of the day, average wind speed of the day, the average pressure of the day, average humidity of the day, dew point, cloud cover, weather cluster, holiday indicator, energy cluster and average load of the preceding day. These features have significant influence on short-term loads which were taken as input features that trained the proposed model to predict the average load value of the day.

Table 1: Weather data

	Avg energy	Max temp	Dew point	Cloud cover	Wind speed	Pressure	Visibility	Humidity	Uv index	Moon phase
Avg energy	1.000000	-0.535188	-0.464499	0.153440	0.105354	-0.064016	-0.150911	0.253010	-0.467917	0.012307
Max temp	-0.535188	1.000000	0.863306	-0.346295	-0.138677	0.098138	0.269981	-0.406919	0.690925	-0.010147
Dew point	-0.464499	0.863306	1.000000	-0.032393	-0.076935	-0.049710	0.048479	0.058871	0.478066	-0.017442
Cloud cover	0.153440	-0.346295	-0.032393	1.000000	0.177538	-0.093291	-0.329287	0.489087	-0.260507	-0.061224
Wind speed	0.105354	-0.138677	-0.076935	0.177538	1.000000	-0.308825	0.286008	-0.035645	-0.127803	0.005563
Pressure	-0.064016	0.098138	-0.049710	-0.093291	-0.308825	1.000000	-0.006675	-0.260783	0.067579	0.009226
Visibility	-0.150911	0.269981	0.048479	-0.329287	0.286008	-0.006675	1.000000	-0.585029	0.256727	0.070642
Humidity	0.253010	-0.406919	0.058871	0.489087	-0.035645	-0.260783	-0.585029	1.000000	-0.540002	-0.009321
Uv index	-0.467917	0.690925	0.478066	-0.260507	-0.127803	0.067579	0.256727	-0.540002	1.000000	-0.009252
Moon phase	0.012307	-0.010147	-0.017442	-0.061224	0.005563	0.009226	0.070642	-0.009321	-0.009252	1.000000

Table 2: Summary of selected features of proposed model

Input features	1. Maximum temperature of the day 2. Average wind speed of the day 3. Average pressure of the day 4. Average humidity of the day 5. Dew point 6. Cloud cover 7. Weather cluster 8. Holiday indicator 9. Energy cluster 10. Average load value of the preceding day
Output feature	1. Average Load value of the day

Performance of the proposed short term load forecasting model

The performance of the proposed model was further compared with four other models as shown in Table 3, the proposed model (Model 1) had a MAPE (%) of 0.31 and RMSE of 1.014. Whereas the DFNN model based on Input Attention Mechanism (IAM) and Hidden Connection Mechanism (HCM), Model 2 has a MAPE (%) of 3.17; model based on Sequence-to-Sequence Recurrent Neural Network (S2S RNN) with Attention, Model 3 had a MAPE (%) of 2.7 and DFNN with sample weights model, Model 4, had 3.22 as RMSE. Model 5, based on DRNN with Levenberg–Marquardt (LM) backpropagation algorithm, had a MAPE of 0.58. The proposed model has significantly outperformed other models with better MAPE and RMSE values due to the filtering technique employed and better learning rate of a deep recurrent neural network as presented in Table 3. The accuracy of the STLF model is improved and eventually promotes the energy demand and supply equilibrium. With better prediction accuracy, the power utilities can make improved load forecasting predictions thereby minimizing the error between the actual supply and forecasted supply.

Table 3: Evaluation results

	Model 1	Model 2	Model 3	Model 4	Model 5
MAPE	0.0031	0.0317	0.027	-	0.058
MAPE (%)	0.31	3.17	2.7	-	-
RMSE	1.014	-	-	3.22	-

Conclusion

The study clearly improved the accuracy of the proposed short-term energy load forecasting model by incorporating Deep Recurrent Neural Networks (DRNN) and sample filtering which provided an exhaustive elucidation for modelling a sophisticated stochastic relationship between the input and output features. Samples were filtered and clustered into several groups according to their degree of similarity and then the model was trained based on those samples and evaluated against other models understudy, namely the Deep Feedforward Neural Network (DFNN) model based on the Input Attention Mechanism (IAM) and Hidden Connection Mechanism (HCM); model based on Sequence-to-Sequence Recurrent Neural Network (S2S RNN) with Attention; model based on DFNN with sample weights and a model based on DRNN with Levenberg-Marquardt (LM) backpropagation algorithm. The proposed model has significantly outperformed other models with mean absolute percentage error and root mean squared error of 0.31% and 1.014, respectively. With such reduction in error, the energy demand and supply chain equilibrium are enhanced, thereby optimizing power deployment and reducing power losses. The proposed model could enhance the efficient energy management which leads to big savings with great economic and environmental benefits.

Conflict of Interest: The authors declare that there are no conflicts of interest.

References

Behnam F, Manar A, Nizar B, and Ursula E 2021 On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN Approach. *IEEE Access* 9: 31191-31212.

Cai Q, Yan B, Su B, Liu S, Xiang M, Wen Y and Feng N 2020 Short-term load forecasting method based on a deep neural network with sample weights. *Int. Trans. Electrical Energy Syst.* 30(5): e12340.

Groß A, Lenders A, Schwenker F, Braun DA and Fischer D 2021 Comparison of short-term electrical load forecasting methods for different building types. *Energy Inform.* 4: 13.

Han HG, Zhang L, Hou Y and Qiao JF 2015 Nonlinear model predictive control based on a self-organizing recurrent neural network. *IEEE Trans. Neural Netw. Learn*

Syst. 27: 402-415.
Huang L, Yang Y, Zhao H, Wang X and Zheng H 2017 Time series modeling and filtering method of electric power load stochastic noise. *Prot. Control Mod. Power Syst.* 2: 25.
Hui X, Qun W, Yao L, Yingbin Z, Lei S and Zhisheng Z 2017 Short-term load forecasting model based on deep neural network. *International Conference on Power and Renewable Energy (ICPRE)* 2: 589-591.
Marinakis V 2020 Big data for energy management and energy-efficient buildings. *Energies* 13(7): 1555.
Mohammad F and Kim YC 2020 Energy load forecasting model based on deep neural networks for smart grids. *Int. J. Syst.Assur. Eng. Manag*. 11(4): 824-834.
Mujeeb S, Javaid N, Javaid S, Rafique A and Ilahi M 2018 Big data analytics for load forecasting in smart grids: A survey. *International Conference on Broadband and Wireless Computing, Communication and Applications.* 10: 77-87.
Narayan A and Hipel KW 2017 Long short-term memory networks for short-term electric load forecasting. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* 2573- 2578.
Papadopoulos S and Karakatsanis I 2015 Short-term electricity load forecasting using time series and ensemble learning methods. *IEEE Power and Energy Conference* 1-6.
Pramono SH, Rohmatillah M, Maulana E, Hasanah RN and Hario F 2019 Deep learning-based short-term load forecasting for supporting demand response program in hybrid energy system. *Energies* 12(17): 3359.
Qiu X, Ren Y, Suganthan PN and Amaratunga GAJ 2017 Empirical mode decomposition-based ensemble deep learning for load demand time series forecasting. *Appl. Soft Comput.* 54: 246-255.
Raza MQ and Khosravi A 2015 A review on artificial intelligence-based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* 50: 1352-1372.
Syed D, Refaat SS and Abu-Rub H 2020 Performance evaluation of distributed machine learning for load forecasting in smart grids. *Cybernetics & Informatics (K&I)* 1-6.
Wei N, Li C, Duan J, Liu J and Zeng F 2019 Daily natural gas load forecasting based on a hybrid deep learning model. *Energimohammes* 12(2): 218.
Xu Q, Yang X and Huang X 2020 Ensemble residual networks for short-term load forecasting. *IEEE Access* 8: 64750-64759.
Zhang M, Yu Z and Xu Z 2020 Short-term load forecasting using recurrent neural networks with input attention mechanism and hidden connection mechanism. *IEEE Access* 8: 186514-186529.
Zhang Q, Lu J, Yang Z and Tu M 2019 A deep learning-based real-time load forecasting method in electricity spot market. *J. Phys.: Conf. Ser.* 1176: 062068.
Zhang Y, Yang R, Zhang J, Weng Y and Hodge BM 2018 Predictive analytics for comprehensive energy systems state estimation. *Big Data Application in Power Systems* 343-376.