Air Compressor Load Forecasting using Artificial Neural Network

Da-Chun Wu*, Babak Bahrami Asl*, Ali Razban**, Jie Chen

Da-Chun Wu, Department of Mechanical Engineering, Purdue University, West Lafayette, Indiana, United States, wudach@iu.edu
Babak Bahrami Asl, Department of Mechanical and Energy Engineering, Purdue University, Indianapolis, Indiana, United States, bahramib@iu.edu
Ali Razban, Department of Mechanical and Energy Engineering, Indiana University-Purdue University, Indianapolis, Indiana, United States, arazban@iupui.edu
Jie Chen, Department of Mechanical and Energy Engineering, Indiana University-Purdue University, Indianapolis, Indiana, United States, jchen3@iupui.edu

* *These authors contributed equally to this work.*
***Corresponding author*. Tel.: +1-317-274-8458; fax: +1-317-274-9744. Dept. of Mechanical & Energy Engineering, 723 West Michigan Street, Indianapolis, IN 46202,

This is the author's manuscript of the article published in final edited form as:

Wu, D. Bahrami, B., Razban, A. and Chen, J., Air Compressor Load Forecasting using Artificial Neural Network, J. Expert Systems with Applications, Vol.168, 15 April 2021, 114209, https://doi.org/10.1016/j.eswa.2020.114209.

Abstract— **Air compressor systems are responsible for approximately 10% of the electricity consumed in United States and European Union industry.** As many researches have proven the effectiveness of using Artificial Neural Network in air compressor performance prediction, there is still a need to forecast the air compressor electrical load profile. **The objective of this study is to predict compressed air systems' electrical load profile, which is valuable to industry practitioners as well as software providers in developing better practice and tools for load management and look-ahead scheduling programs. Two artificial neural networks, Two-Layer Feed-Forward Neural Network and Long Short-Term Memory were used to predict an air compressors electrical load. Compressors with three different control mechanisms are evaluated with a total number of 11,874 observations. The forecasts were validated using out-of-sample datasets with 5-fold cross-validation. Models produced average coefficient of determination values from 0.24-0.94, average root-mean-square errors from 0.05 kW - 5.83 kW, and mean absolute scaled errors from 0.20 - 1.33. The results indicate that both artificial neural networks yield good results for compressors using variable speed drive (average R^2 = 0.8 and no naïve forecasting), only the long short-term memory model gives acceptable results for compressors using on/off control (average R^2 = 0.82 and no naïve forecasting), and no satisfactory results are obtained for load/unload type air compressors (models constituting naïve forecasting).**

Keywords—Load forecasting, Air compressor, Artificial Neural Network, FFNN, LSTM

1. Introduction

Energy management and its importance are receiving increased attention due to the shortage of fossil fuels, stricter environmental regulations, and increasing utility prices. In past, energy costs were viewed merely as overhead instead of as an important cost function. This behavior has changed during the last decade. As a result, more companies and organizations have started to develop energy management plans and to improve energy efficiency using data collection and analysis (Schulze et al., 2016).

To improve energy efficiency, using historical data for future electrical load prediction is vital for both utility providers and users. Decisions such as power system operation, maintenance, and planning can be made based on an accurate load forecasting method (Kamel & Baharudin, 2007). An accurate energy forecasting model is essential for each subsystem in an industrial facility to be able to operate ideally in regard to the factory's overall energy usage. By introducing sensors and advanced control systems, the forecasting algorithms can also be connected to react to different scenarios and optimally operate each subsystem with the ultimate goal of optimizing the total factory energy usage to minimize cost and increase energy efficiency. The main burden in this kind of prediction is that there are multiple parameters that need to be taken into consideration and high fluctuation in subsystem demand patterns.

The compressed air system is one of the significant energy users in industrial plants. This system is responsible for approximately 10% of the electricity consumed in the industry in the United States and the European Union (Saidur et al., 2010; Zahlan & Asfour, 2015). Predicting the compressed air system's electrical load will be valuable to both industry practitioners and related software providers in developing better practices and tools for load management and look-ahead scheduling programs.

Based on the required time ahead for decision-making, there are different categories of load forecasting: short-term load forecasting (STLF) with a range of a few minutes ahead up to a day ahead (Platon et al., 2015), medium-term load forecasting (MTLF) with a range of one week to a year ahead (De Felice et al., 2015) and long-term load forecasting (LTLF) from one year to decades (Kandil et al., 2002). The focus of this paper is STLF.

Various methodologies have been utilized for STLF which can be divided into two classifications: conventional statistical methods and artificial intelligence (AI) methods. Some statistical methods have been proposed, which include regression models (Vu et al., 2015), autoregressive integrated moving average (ARIMA) model with a lifting scheme (Box et al., 2008; Lee & Ko, 2011), and adaptive-rate-of-change (ARC) algorithm model (D. C. Wu et al., 2018). Machine learning techniques such as artificial neural networks (ANN) with backpropagation (Werbos, 1990), support-vector network (Cortes & Vapnik, 1995), and particle swarm optimization (Kennedy & Eberhart, 1995) are considered AI methods. In the study of electrical load forecasting, researchers have investigated neural networks combined with a fuzzy logic algorithm (Dash, 1995; Zadeh, 1988), support vector machine and ant colony optimization (Niu et al., 2010), and particle swarm optimization with Gaussian and adaptive mutation (Q. Wu, 2010).

For forecasting problems, ANNs possess several advantageous traits including universal approximation, data driven format, and nonlinear time series capability. (Abbasimehr et al., 2020). They are also more reliable when applied touncertain scenarios and are not reliant on human expertise. In general, forecasting with ANNs is superior to forecasting with statistical methods. Faster convergence speed of the networks, lower computational complexity, reduced training period, better generalization and enhanced network performance can be gained by increasing correlation impact, pre-processing the training data, and utilizing optimal network structure and better learning algorithms (Shekhar & Amin, 1992).

In this research, we consider two matured ANNs, Two-Layer Feed-Forward Neural Network (FFNN) with time delays and Long Short-Term Memory (LSTM) recurrent neural network, to predict the air compressor electrical load. The contributions of this paper are:

1. a survey of two ANNs in the application of air compressors electrical load forecasting by predicting day-ahead electrical power.
2. use of the Hampel filter in pre-processing to assist in outlier elimination.

The rest of this paper is organized as follows: Section 2 presents previous studies in electrical load forecasting. A brief overview of FFNN and LSTM is given in Section 3. Section 4

* *These authors contributed equally to this work.*
***Corresponding author.* Tel.: +1-317-274-8458; fax: +1-317-274-9744. Dept. of Mechanical & Energy Engineering, 723 West Michigan Street, Indianapolis, IN 46202,

describes the experiment process and the results. Section 5 discusses the results. The paper concludes in Section 6.

2. Related works

2.1. ANN in forecasting

ANNs are widely utilized in demand forecasting due to the high complexity of energy systems and ANNs' ability to simplify the non-linear modeling (Deb et al., 2016). Neto and Fiorelli (2008) and Wong et al. (2010) performed building energy consumption forecasting utilizing both ANN and EnergyPlus. In each study, ANN outperforms EnergyPlus in prediction precision. The ANN model used by Kalogirou and Bojic (2000) to predict the energy consumption of a passive solar building gave coefficient of determination (R^2) of 0.9991 when completely unknown data were presented to the network. In their research, Aydinalp et al. (2004) used ANN and simulation to model space heating and domestic hot water energy usage. The ANN model for domestic hot water forecasting had R^2 of 0.871 while the simulation had R^2 of 0.828. The ANN model for space heating forecasting had R^2 of 0.908 while the engineering model had R^2 of 0.778. Hybrid ANN models involving LSTM were also developed for wind speed forecasting (Memarzadeh & Keynia, 2020; Rodrigues Moreno et al., 2020). Abbasimehr et al. (2020) proposed a multilayer LSTM network for demand forecasting, resulting in a symmetric mean absolute percentage error of 0.1085. Ribeiro et al. (2019) used wavelet neural networks to predict hourly electrical load data obtained from Italy and the Global Energy Forecasting Competition, 2012. The results showed R^2 around 0.75 in the former study and 0.92 in the later when predicting load values 24 hours in advance. Raza and Khosravi (2015) had a comprehensive literature review regarding the AI-based load demand forecasting techniques and presented a performance evaluation of AI techniques and their current potential. The results showed an average mean absolute percentage error (MAPE) of 2.57% when applying AANs for STLF.

2.2. ANN usage in air compressor research

Improving compressors' performance is always the main goal of designers. Compressor performance profiles have great importance for simulation and design purposes. Experimental values at various operating conditions are being used to draw these maps. These experiments are time-consuming, costly and cannot predict future performance. Therefore, many researchers have used machine learning techniques to predict performance curves for different operating conditions. Fei et al. (2016) used a novel artificial neural network integrating the feed-forward back-propagation neural network with Gaussian kernel function. The result showed an 80% agreement with the experimental data. Yu et al. (2007) applied a three-layer back-propagation neural-network with the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) to predict stage-by-stage axial-compressor performance. General regression neural network (GRNN) (Specht, 1991), rotated general regression neural network (RGRNN) (Gholamrezaei & Ghorbanian, 2007), radial basis function network (RBFN) (Broomhead and Lowe, 1988), and multilayer perceptron (MLP) network were applied by Ghorbanian and Gholamrezaei (2009) to predict the compressor performance. Tian et al. (2015) used hybrid ANN with partial least squares (PLS) model to predict the thermodynamic performance of a scroll compressor and the model had the R^2 of 0.9999 when compared to the experimental data. A variable-speed reciprocating compressor had been studied to predict mass flow rate, power consumption and discharge temperature at different operating conditions by using ANN (Ledesma et al., 2015). Consequently, compressor performance prediction using ANN has been proven useful, but there is still a need to forecast the air compressor electrical load profile to assist STLF because air compressors are large energy consumers and typical Significant Energy Uses (SEU) in the manufacturing industry.

3. Neural network

ANNs are machine learning tools that process data similar to the human brain. ANNs can construct linear and nonlinear models for time series. They are widely accepted as effective tools to forecast difficult time-series problems. We select two types of ANNs to construct the models: One is FFNN with time delays, and the other one is sequence to sequence (S2S) LSTM. The forecasting capability of FFNN has been widely tested on many time series problems (Bhaskar & Singh, 2012; Firat et al., 2010; Hu et al., 2001; Yona et al., 2007). While FFNN only allows signals to travel one way -- as if there is no feedback loop -- LSTM can solve many time series tasks unsolvable by feed-forward networks and has shown outstanding performance in predicting electrical load (Atef & Eltawil, 2020; Gers et al., 2002). In this section, a brief description of FFNN and LSTM is presented.

3.1. Feed-forward neural network

One of the concentrations of this study is on FFNN because FFNN is one of the most commonly used ANNs for energy forecasting (Jovanović et al., 2015). The main task of these networks is to approximate unknown relationships between inputs and outputs.

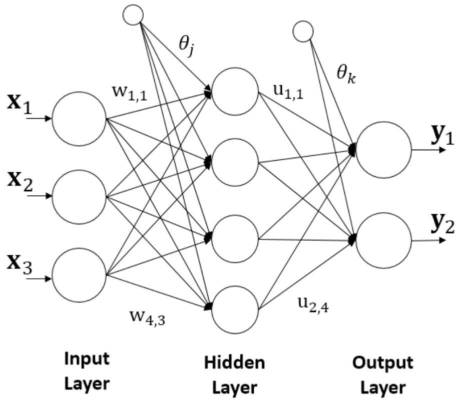

Fig. 1. A three-layer feed-forward neural network.

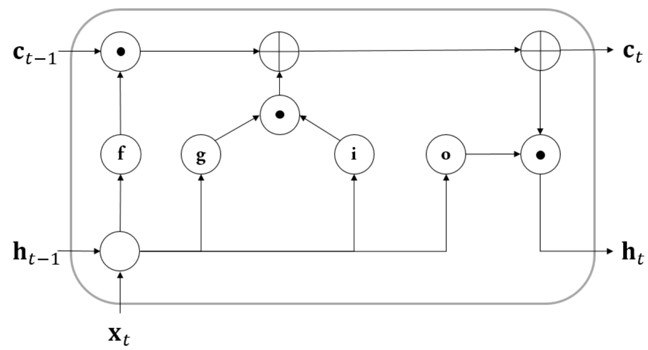

Fig. 2. A three-layer feed-forward neural network.

Typically, two steps are required for processing: the first step is the linear combination of input data; the second step is using the first step's outcome as an input to a nonlinear activation function. The combination exercises the weights associated with each link and a fixed bias term θ (Hippert et al., 2001).

Feed-forward neural networks consist of layers of processing units denoted as neurons. Neurons are arranged by the layers -- input/output layers and hidden layers. Input layers read in a signal to the network while hidden layers pass the signal through the network and weighted connections. Every hidden neuron receives input in the form of weighted signals from the previous layer. These weighted signals then flow to the output layer in one direction as shown in a three-layer FFNN in Fig. 1 (Haykin, 2007). While there is no connection between neurons in the same layer, the number of input neurons correlates to the number of input layers, and the number of output neurons correlates to the number of output layers. The user chooses the number of hidden layers and the number of neurons in each hidden layer. In this network, the weight matrices are $W_{i,j}$ and $u_{j,k}$. The bias vector which can be applied to inputs, outputs, or both is annotated with θ. By using identity function ($y = x$) as an example for the activation of the hidden layer and the linear functions for the output layer, the network output is described as Eq. (1) (Grolinger et al., 2016; Hippert et al., 2001):

$$y_k = \sum_{j=1}^{4}(u_{jk}.\sum_{i=1}^{3}(w_{ij}x_i + \theta_j)) + \theta_k \tag{1}$$

According to (1), the output computation of a simple network can become complicated. FFNN weights are estimated during the training phase. Many optimization algorithms such as back-propagation (Werbos, 1990) with gradient descent optimization have been developed to calculate the weights. Back-propagation utilizes the steepest-descent technique based on the calculation of the gradient of the loss function for the network. The learning process starts with assigning random numbers to the weights. Afterward, the calculated output based on the input will be used to compare the network output to the actual output and adjust the weights accordingly until we have a network that has an output with an acceptable predefined error threshold (Grolinger et al., 2016; Hippert et al., 2001).

3.2. Long short-term memory

While FFNN has limitations, LSTM can solve many time series tasks unsolvable by feed-forward networks (Gers et al., 2002). LSTM is a special type of recurrent neural networks. A typical LSTM unit is made of a cell, an input gate, an output gate and a forget gate as shown in Fig. 2. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. These gate operations are expressed in the following equations from Eq. (2) to Eq. (8) (Hochreiter & Schmidhuber, 1997),

$$i_t = \sigma(W_i x_t + R_i h_{t-1} + b_i) \tag{2}$$

$$f_t = \sigma(W_f x_t + R_f h_{t-1} + b_f) \tag{3}$$

$$g_t = \phi(W_g x_t + R_g h_{t-1} + b_g) \tag{4}$$

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + b_o) \tag{5}$$

$$\mathbf{W} = [W_i, W_f, W_g, W_o] \tag{6}$$

$$\mathbf{R} = [R_i, R_{f,} R_g, R_o] \tag{7}$$

$$\mathbf{b} = [b_i, b_f, b_{g,} b_o] \tag{8}$$

where i_t, f_t, and o_t corresponds to the input gate, the forget gate, and the output gate, respectively. g_t is the added cell information to state. $\mathbf{W}$, $\mathbf{R}$, and $\mathbf{b}$ are learnable weights pertaining to the input, the recurrent, and the bias, respectively. σ and ϕ are sigmoid activation and hyperbolic tangent, respectively. Also, the inputs and the outputs are defined as follows from Eq. (9) to Eq. (11) (Hochreiter and Schmidhuber, 1997),

$$\mathbf{x}_t = [x_{1t}, x_{2t}, \ldots, x_{Ct}] \tag{9}$$

$$\mathbf{c}_t = f_t \odot \mathbf{c}_{t-1} + i_t \odot g_t \tag{10}$$

$$\mathbf{h}_t = o_t \odot \phi(\mathbf{c}_t) \tag{11}$$

where $\mathbf{x}_t$, $\mathbf{h}_t$, and $\mathbf{c}_t$ are arrays corresponding to the input time series with C features, the cell state, and the output at time step t. $\odot$ denotes the Hadamard product.

Being capable of unknown duration lagging between important events in a time series makes LSTM networks capable of organizing, processing and making predictions based on time series data (Shi et al., 2015).

Table 1
General description of the dataset.

	Compressor 1			Compressor 2			Compressor 3		
	N	Mean	Std.dev.	N	Mean	Std.dev.	N	Mean	Std.dev.
Active Power (kW)	3603	26.05	6.97	2892	28.55	23.97	5379	8.83	0.63
Inlet Temperature (deg C)	3603	24.80	2.35	2892	30.45	3.93	5379	26.10	0.31
Line Pressure (bar)	3603	8.45	0.26	2892	2.51	0.38	5379	4.37	0.08

4. Model design and implementation

This section first introduces three datasets used in this study and then proceeds with data preprocessing to describe the method to eliminate outliers. Finally, the section discusses the proposed neural networks with high-level architecture.

4.1. Data collection

Data were collected from three types of air compressors and their surroundings. The dataset contains the active power, the temperature of the air inlet, the line pressure, and the timestamp for the individual air compressor. The sampling rate was 15 minutes. In practice, the electrical load is measured every 15 minutes in most of the industrial domain (Nolde and Morari, 2010). The data is presented in Table 1.

Compressor 1 utilizes variable frequency drives (VFD) to adjust the speed of the compressor motor. Compressor 2 has a load/unload with an automatic shutdown that allows itself to cycle between full-load condition and low-load condition, where a compressor consumes significantly less energy in the latter than in the former. Compressor 3 shuts itself down when the pressure in the air tank reaches the set pressure and restarts once the pressure drops below the setting. Table 2 summarizes the system specification, which includes the control type, compressors' rated power, compressors' full-load current, the maximum discharge pressure, the air receiver size, and the set pressure.

Table 2
System specifications of Compressor 1, 2 and 3.

System	Parameter	Value	Unit
Compressor 1 (VFD)			
	Compressor motor capacity	37	kW
	Full-load current	71	A
	Maximum discharge pressure	13	bar
	Air dryer rated power	2.7	hp
	Air receiver size	500	liter
Compressor 2 (load/unload w/ auto shutdown)			
	Compressor motor capacity	74.57	kW
	Full-load current	96	A
	Maximum discharge pressure	12	bar
	Air dryer rated power	1.7	hp
	Air receiver size	1,500	liter
Compressor 3 (on/off)			
	Compressor motor capacity	37	kW
	Full-load current	71	A
	Maximum discharge pressure	13	bar
	Air dryer rated power	2.7	hp
	Air receiver size	500	liter

4.2. Data analysis and preprocessing

It is not uncommon to have outliers and erroneous values in the data collected by different sensors; therefore it is important to pre-process the inputs before the neural network can be built (Fawzy et al., 2013). Shown in Fig. 3, Compressors 1 and 3 have several outliers greater than three standard deviations from their respective means.

Because of the outliers, the Hampel filter was used to eliminate the outliers from the electrical power dataset. The Hampel filter (Hampel, 1974) is a variation of the three-sigma rule of statistics that is robust against outliers (Liu et al., 2004). The Hampel filter replaced the outliers with a local median, keeping the length of the dataset array constant, which made real-time machine learning algorithms feasible. A brief explanation of the Hampel filter is as follows:

Given a sequence $x_1, x_2, x_3, \dots, x_n$ and a sliding window half-width, K, we can define Eq. (12) and Eq. (13) (Pearson et al., 2016),

$$W_i^K = (x_{i-K}, \dots, x_i, \dots, x_{i+K}) \tag{12}$$

$$m_i = median(x_{i-K}, \dots, x_i, \dots, x_{i+K}) \tag{13}$$

where W_i^K is a set of numbers within a moving window, and m_i is the median value from the moving window.

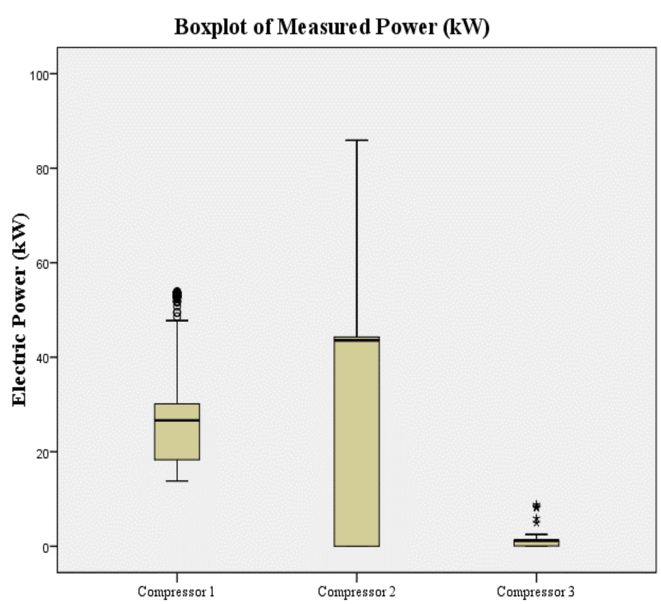

Fig. 3. Boxplot showing measured electrical power of three air compressors.

Table 3
Descriptive statistics of electrical power data after preprocessing.

	Compressor 1					Compressor 2					Compressor 3				
Mode	N	Min	Max	Mean	Std. Dev.	N	Min	Max	Mean	Std. Dev.	N	Min	Max	Mean	Std. Dev.
Working	2,649	18.53	50.89	33.96	4.53	2,124	0.01	85.9	28.55	23.97	5,379	0.05	1.37	0.82	0.58
Non-working	949	20.5	22.7	21.48	0.32	768	0.01	85.19	11.47	20.13	-	-	-	-	-

After applying the Hampel filter to the sequence, a new dataset of responses was obtained using Eq. (14) (Pearson et al., 2016),

$$y_i = \begin{cases} x_i, & |x_i - m_i| \le tS_i \\ m_i, & |x_i - m_i| > tS_i \end{cases} \tag{14}$$

where t is a positive integer and S_i is the median absolute deviation (MAD) defined as Eq. (15) (Pearson et al., 2016).

$$S_i = 1.4826 \times median(|x_{i-K} - m_i|, \ldots, |x_{i+K} - m_i|) \tag{15}$$

Outliers were replaced by moving medians after choosing the Hampel filter with $K = 5$ and $t = 2$, which represented a 75-minute sliding window and two MAD as suggested by Pearson et al. (2016).

An air compressor's electrical load is greatly affected by the production schedule. The temporal variation caused by weekdays and holidays affects the outcomes of the models. Fig. 4 shows the average kW consumed by each compressor as a function of days.

From Fig. 4, Compressor 1 and Compressor 2 have a main working schedule from Monday to Friday. Compressor 3 has a constant working pattern. As a result, it is reasonable to divide electrical power data of Compressor 1 and 2 into two categories: 1) working days and 2) non-working days for further analysis. After preprocessing the data, the results are shown in Table 3.

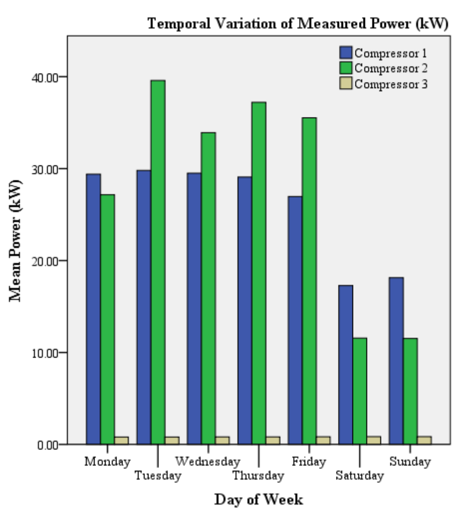

Fig. 4. Temporal variation of the measured power of the three compressors.

4.3. Neural network deployment

For FFNN, each output was predicted by a vector of input including the time of day, the day of week, the pressure in the compressed air line, the temperature of the intake air, and the historical kW data in the 15-minutes interval. To determine which historical kW data is valuable, we performed autocorrelation (AC) and partial autocorrelation (PAC) on the dataset (Box et al., 2008). Fig. 5 shows an example of the autocorrelation analysis done on Compressor 1. The blue lines indicate a 99% confidence bound. It can be seen from the figure that although AC shows high autocorrelation between Lag 1 and Lag 31, only Lag 1 to Lag 3 are significant when effects from previous lags are removed, as indicated in the PAC plot.

According to Hecht-Nielsen (1989), the number of hidden layer neurons is recommended by the following relationship as $m = 2n + 1$, where m is the number of hidden layer neurons and n is the number of input neurons. The architecture of FFNN is presented in Fig. 6.

For LSTM, only the historical kW data was used to predict the target output. Fig. 7 shows the architecture of S2S LSTM used to forecast electrical load, which utilizes all the historical kW data.

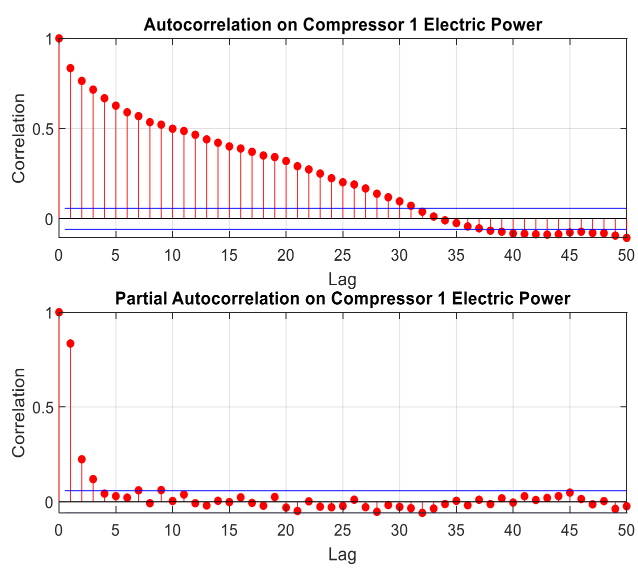

Fig. 5. ACF and PACF analysis on compressor 1 electrical power.

To prevent neural networks from overfitting, the model requires a large dataset (Srivastava et al., 2014). We utilized all collected data from measurement for the training, leaving only 96 time-steps (1 day) for the testing. We followed the procedure of 5-fold cross-validation by withholding different testing dataset among the 5 trials.

The results were obtained by using MATLAB machine learning toolbox (Kim, 2017), which is designed for solving time-series problems with nonlinear input-output. To perform STLF, we forecasted the electrical load of each compressor 1 time-step (15 minutes) into the future. A summary of the ANN design parameters and inputs are listed in Table 4.

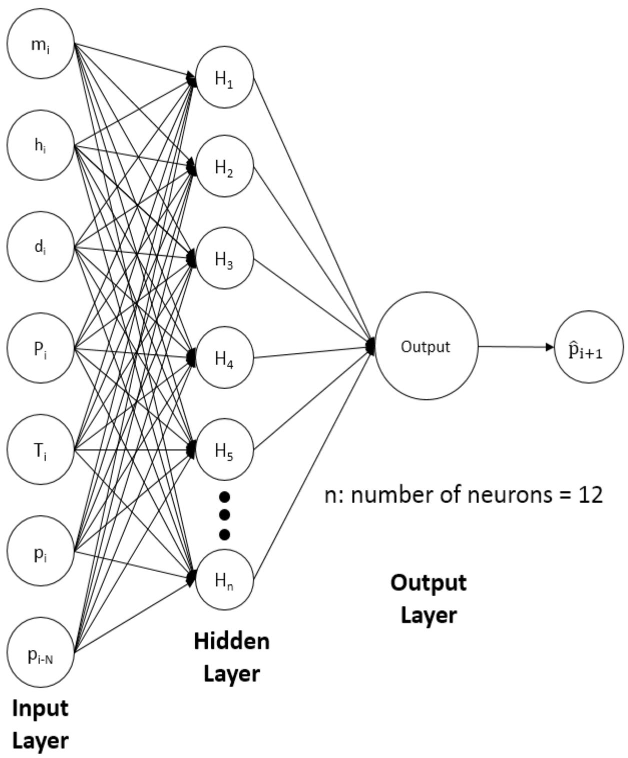

Fig. 6. FFNN architecture used in the study. Inputs include minute (m), hour (h), day of week (d), pressure (P), temperature (T), and historical power (p). The output is the estimated electrical power for the next 15 minutes.

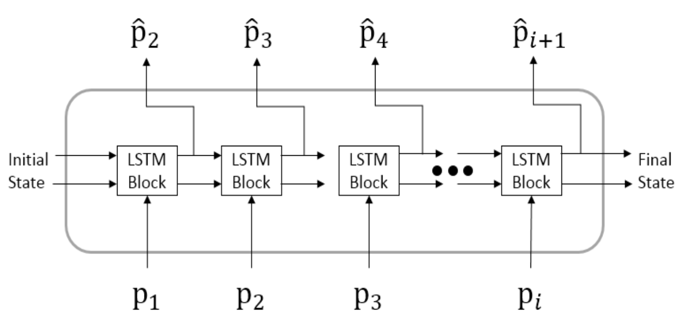

Fig. 7. LSTM architecture used in the study. Inputs include all historial power (p). The output is the estimated electrical power for the next 15 minutes.

Table 4
Summary of ANN design parameters and inputs.

Network type	FFNN	LSTM
Inputs	minute, hour, day of week, pressure (bar), temperature (°C), power (kW)	power (kW)
Targets	power (kW)	power (kW)
Training algorithm/Solver	Levenberg-Marquardt	Adam
Performance	MSE	RSME

4.4. Performance evaluation

We used 2 statistical metrics and 1 comparison metric to evaluate the performance of each forecasting model. The metrics included coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute scaled error (MASE) as presented by Hyndman and Koehler (2006). R^2 is calculated by Eq. (16).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i|^2}{\sum_{i=1}^{n} |Y_i - \bar{Y}|^2} \tag{16}$$

R^2 describes how well the prediction model fits the observation. RMSE calculates the deviation between the values predicted by the model and the observed values using Eq. (17).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} |Y_i - \hat{Y}_i|^2}{n}} \tag{17}$$

Last, to prove that the models are better than the naïve method of forecasting which simply takes the last observation as the forecast, we calculated MASE using Eq. (18).

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n} |Y_i - \hat{Y}_i|}{\frac{1}{m-1}\sum_{j=2}^{m} |Y_j - Y_{j-1}|} \tag{18}$$

Here, Y_i represents the actual measured values; $\hat{Y}_i$ represents the predicted values by the model; $\bar{Y}$ is the mean of the measured sample; n and m are the number of data used for performance evaluation and comparison, respectively.

5. Results and discussion

After the neural networks were trained, we applied them to the remaining 96 time-steps. The results then were evaluated by the performance metrics described in the previous section in Table 5. The measured values and predicted values are shown in Fig. 8 to Fig. 12. Only predictions with the best R^2 are plotted.

The results show mixed performances regarding forecasting ability. The FFNN model performs better when the variation of the data is higher as seen during working days whereas the LSTM model performed better when the fluctuation in the data is smaller and noisier. The factor that might contribute to this issue could be the numbers of input. The FFNN model uses more inputs related to the operating conditions and therefore yields better prediction. However, the LSTM model can take care of more noisy data without many input variables (Weninger et al., 2015). Nevertheless, both models have difficulty catching peak values in the measurement. It is a known issue that also exists in other ANN implementation such as wind power generation forecasting (Mason et al., 2018).

Table 5
Evaluation of models with 5-fold cross-validation and their respective averages

		Compressor 1 (VFD)						Compressor 2 (load/unload w/ auto shutdown)						Compressor 3 (on/off)		
		Working day			Non-working day			Working day			Non-working day			Working day		
Model	Trial	R^2	RMSE (kW)	MASE	R^2	RMSE (kW)	MASE	R^2	RMSE (kW)	MASE	R^2	RMSE (kW)	MASE	R^2	RMSE (kW)	MASE
FFNN	1	0.76	2.56	0.78	0.77	0.07	0.73	0.93	7.58	1.12	0.91	5.77	1.12	0.22	0.48	0.60
	2	0.80	1.64	0.83	0.86	0.05	1.09	0.90	8.53	0.94	0.92	5.18	1.23	0.30	0.49	0.49
	3	0.81	1.64	0.85	0.75	0.05	0.94	0.89	2.89	1.14	0.93	5.58	1.07	0.29	0.49	0.50
	4	0.86	1.33	0.65	0.69	0.07	0.99	0.92	3.41	0.94	0.96	5.54	1.05	0.25	0.50	0.53
	5	0.76	2.02	0.88	0.71	0.06	1.09	0.87	6.74	1.04	0.91	5.75	1.10	0.16	0.54	0.70
	Avg.	0.80	1.84	0.80	0.76	0.06	0.97	0.90	5.83	1.04	0.93	5.56	1.11	0.24	0.50	0.56
LSTM	1	0.89	1.21	0.74	0.74	0.07	0.74	0.87	9.80	1.79	0.94	5.79	0.84	0.71	0.32	0.32
	2	0.82	1.52	0.91	0.84	0.04	0.74	0.84	8.39	1.31	0.92	4.21	1.25	0.82	0.23	0.20
	3	0.81	1.53	0.92	0.89	0.04	0.74	0.94	5.01	1.25	0.95	5.62	1.07	0.86	0.23	0.15
	4	0.72	1.72	1.10	0.82	0.05	0.68	0.84	2.39	0.77	0.93	4.89	0.92	0.84	0.24	0.16
	5	0.91	1.22	0.74	0.80	0.07	0.78	0.96	5.24	1.51	0.96	5.74	0.98	0.85	0.24	0.19
	Avg.	0.83	1.44	0.88	0.82	0.05	0.74	0.89	6.17	1.33	0.94	5.25	1.07	0.82	0.25	0.20

5.1. Load forecasting for Compressor 1 (VFD)

Table 5 shows that the LSTM model gives better average R^2 values and lower RMSE values in forecasting both working and non-working electrical loads. For the forecast of the working day's electrical load, however, the results of both models are comparable. The FFNN model performs better in terms of MASE which suggests that the LSTM model is closer to the naïve forecasting, but the difference is not conclusive. The LSTM model outperforms the FFNN model in forecasting the non-working electrical load in every evaluation metric. Particularly, MASE suggests that FFNN is merely taking the previous datapoint as the forecast of the next value. This finding is consistent with the statement that LSTM can handle more noisy data. Fig. 8 and Fig. 9 show each model's prediction and the measured values.

5.2. Load forecasting for Compressor 2 (load/unload w/ auto shutdown)

Fig. 10 and Fig. 11 show the predictions and measured values. Although the average R^2 values of both models are acceptable, their average MASE values are all above unity, which means that the errors they produce are larger than naïve forecasting. As a result, we cannot conclude the effectiveness of using either FFNN or LSTM under this situation. The reason for these high R^2 values can come from the fact that in both working and non-working days the electrical load profile exhibits a long period of zeros.

5.3. Load forecasting for Compressor 3 (on/off)

Fig. 12 shows the models' prediction. Compressor 3 has a control type of on/off dual control, therefore, the measurement is similar to a binary categorical value. Brouwer (2002) investigated the forecasting ability of FFNN with categorical inputs and found that the result is not sound. This could be one of the reasons that the FFNN model did not yield an acceptable result in this case. On the other hand, the LSTM model returns good results with an average R^2 value over 0.8 and a low MASE value.

In summary, both models yield a decent prediction for air compressors with VFD. The LSTM model has a good forecasting ability for the electrical load of compressors with an on-and-off control. However, both models fail to give an acceptable prediction for the load/unload type air compressors.

6. Conclusions and future work

In this research, we investigated the capability of an artificial neural network to predict the electrical load of three compressors of different control mechanisms. We used the time of day, the day of week, the pressure in the compressed air line, the temperature of the air intake, and historical power values to build the forecasting models. The results show that both models are good for predicting a compressor's electrical load that utilizes VFD, with average R^2 values of 0.80 and 0.83, respectively. LSTM performs better in forecasting the electrical load of a compressor with on/off dual control with an average R^2 value of 0.82, compared to that of FFNN of 0.24. Both models were incapable of providing a good prediction for load/unload type air compressor.

Further research may include:

- Investigation of electric demand peak forecasting using ANNs.
- Investigation of other machine learning technologies for load/unload type air compressor electrical load forecasting.
- Integration of ANN with other learning techniques such as the support vector machine (SVM) to further improve forecast accuracy.

The ability to forecast the electrical power consumption of an air compressor is a huge step towards a successful demand response and smart manufacturing for two main reasons. An air compressor is not only is a typical SEU but also an indispensable component in the manufacturing industry.

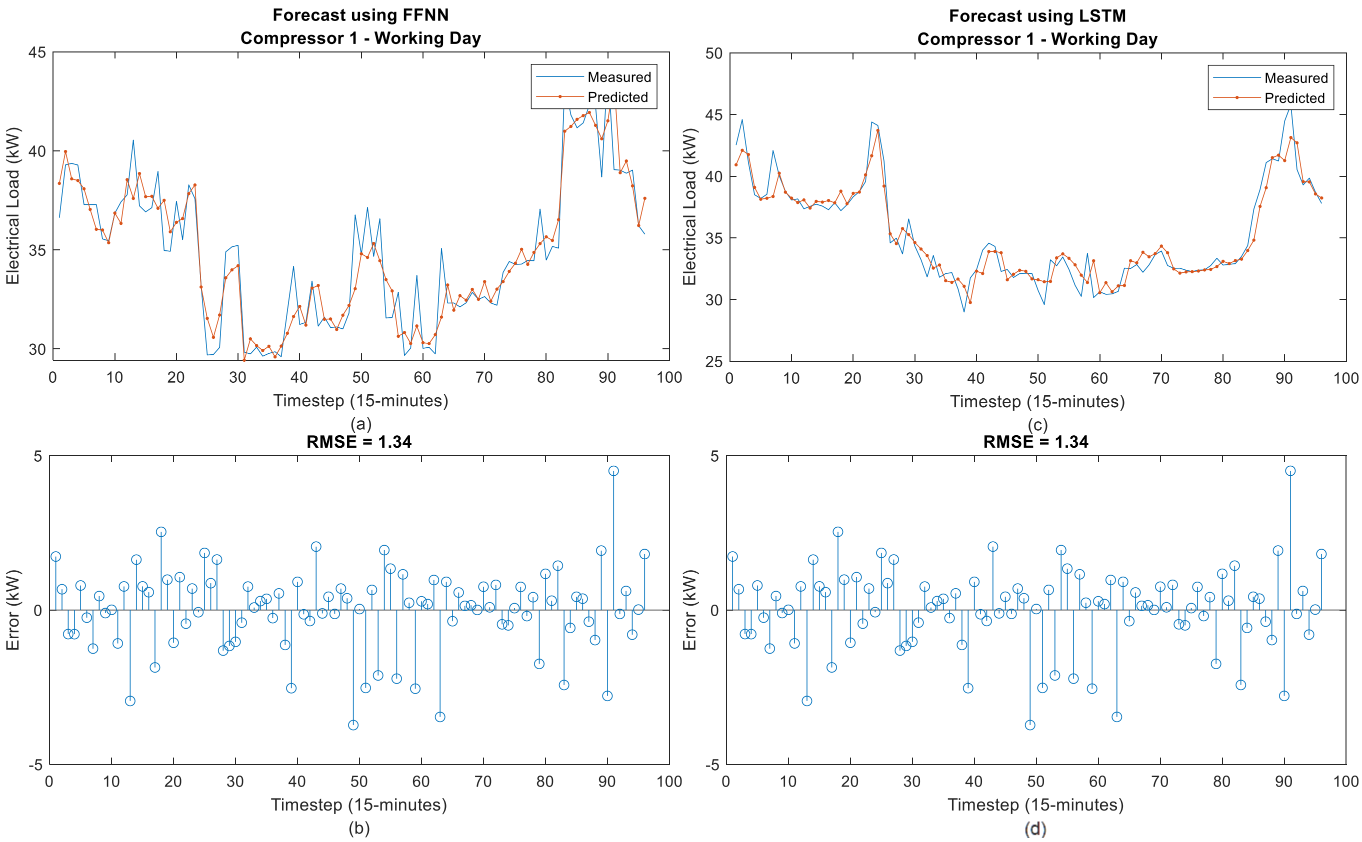

Fig. 8. Best predictions and their respective errors to the measurement of a working day in 15 minutes interval using FFNN and LSTM for Compressor 1.

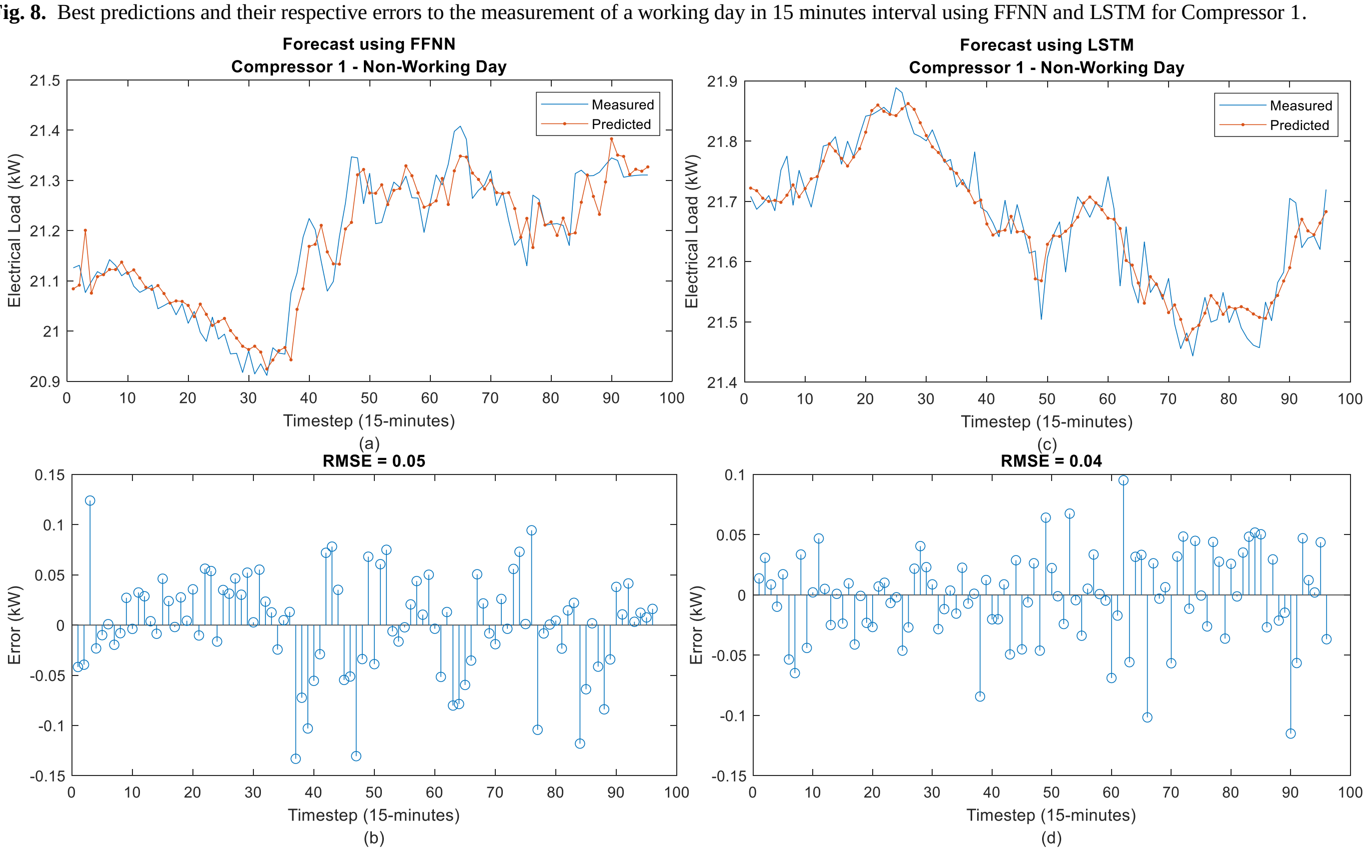

Fig. 9. Best predictions and their respective errors to the measurement of a non-working day in 15 minutes interval using FFNN and LSTM for Compressor 1.

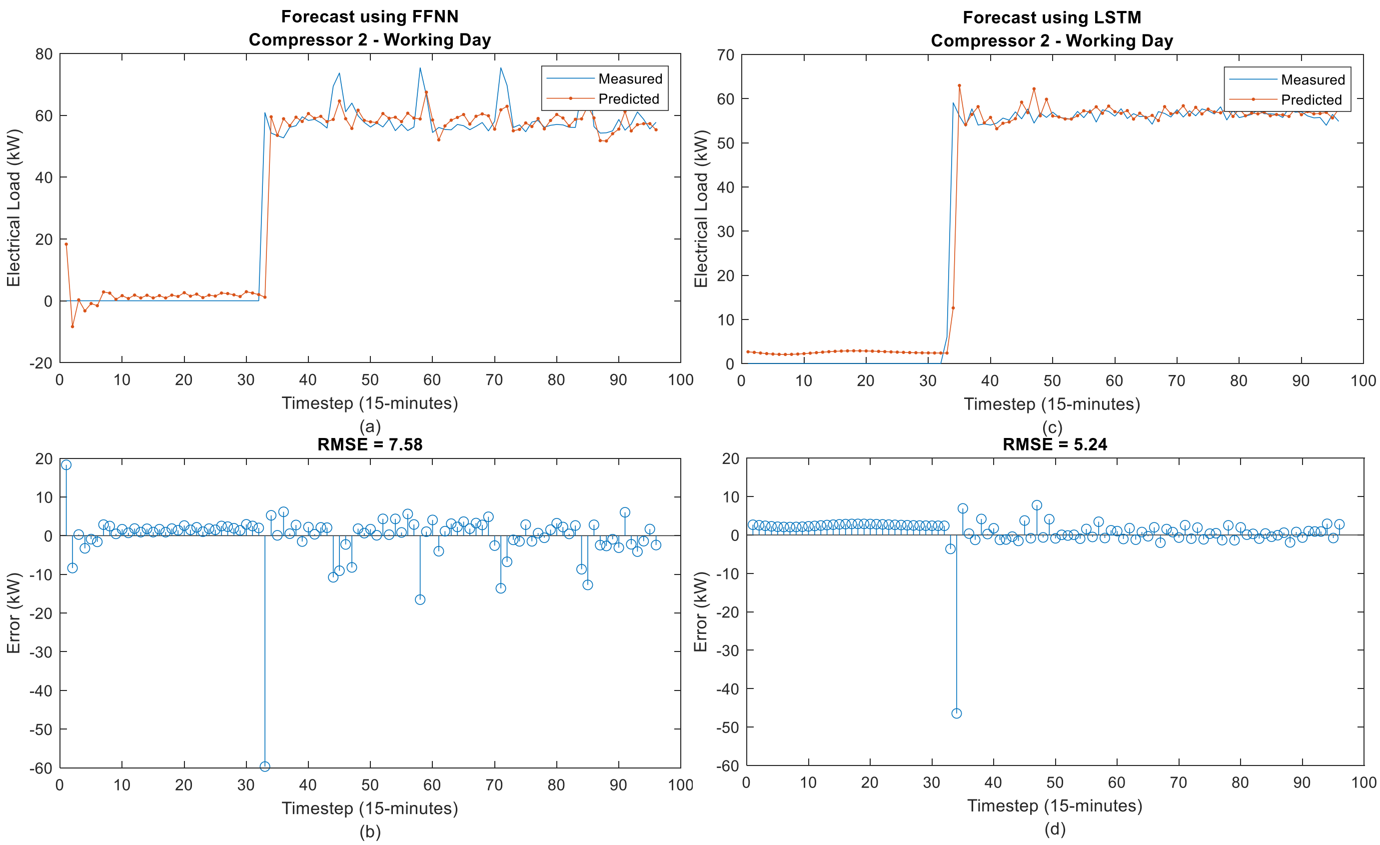

Fig. 10. Best predictions and their respective errors to the measurement of a working day in 15 minutes interval using FFNN and LSTM for Compressor 2.

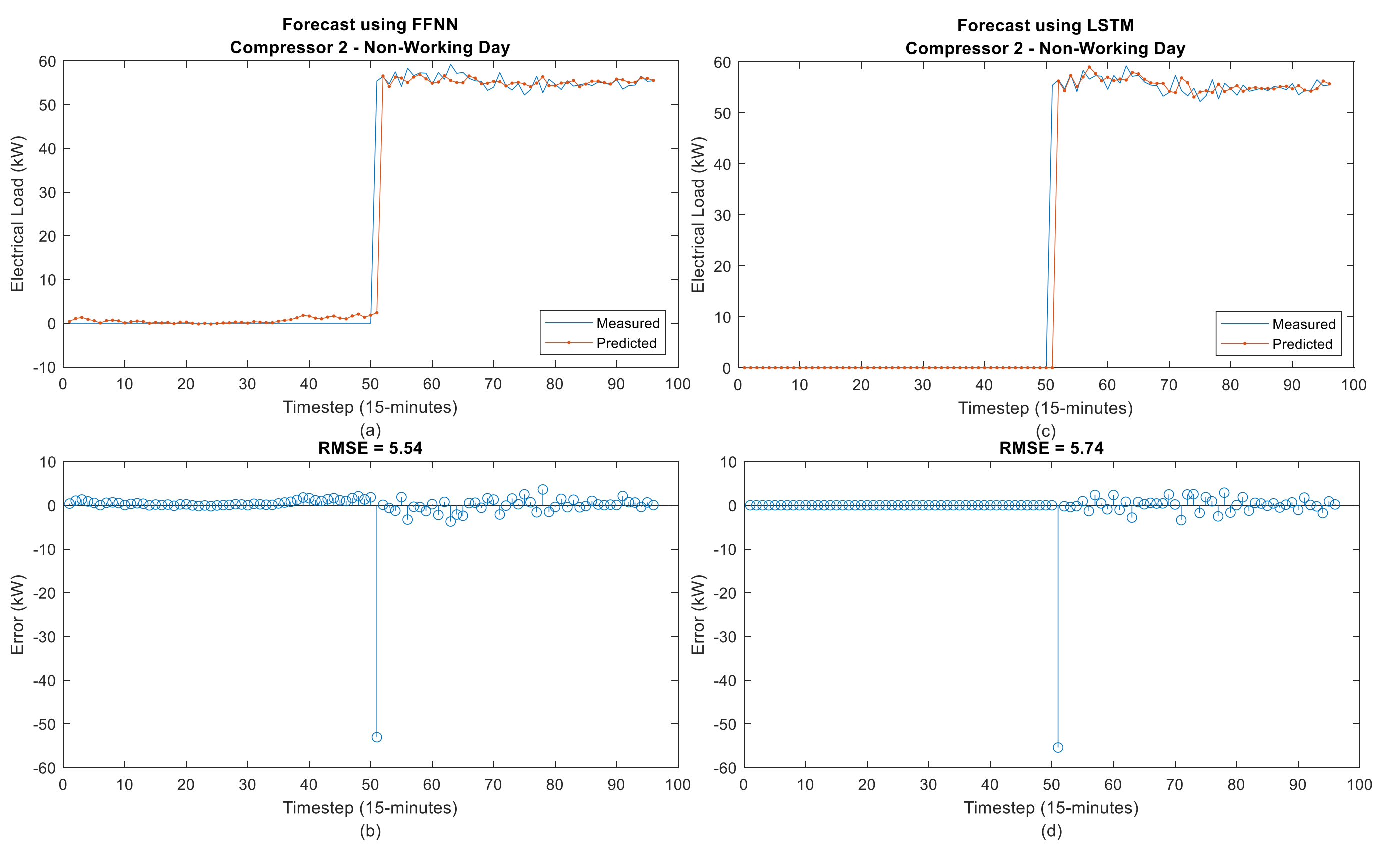

Fig. 11. Best predictions and their respective errors to the measurement of a non-working day in 15 minutes interval using FFNN and LSTM for Compressor 2.

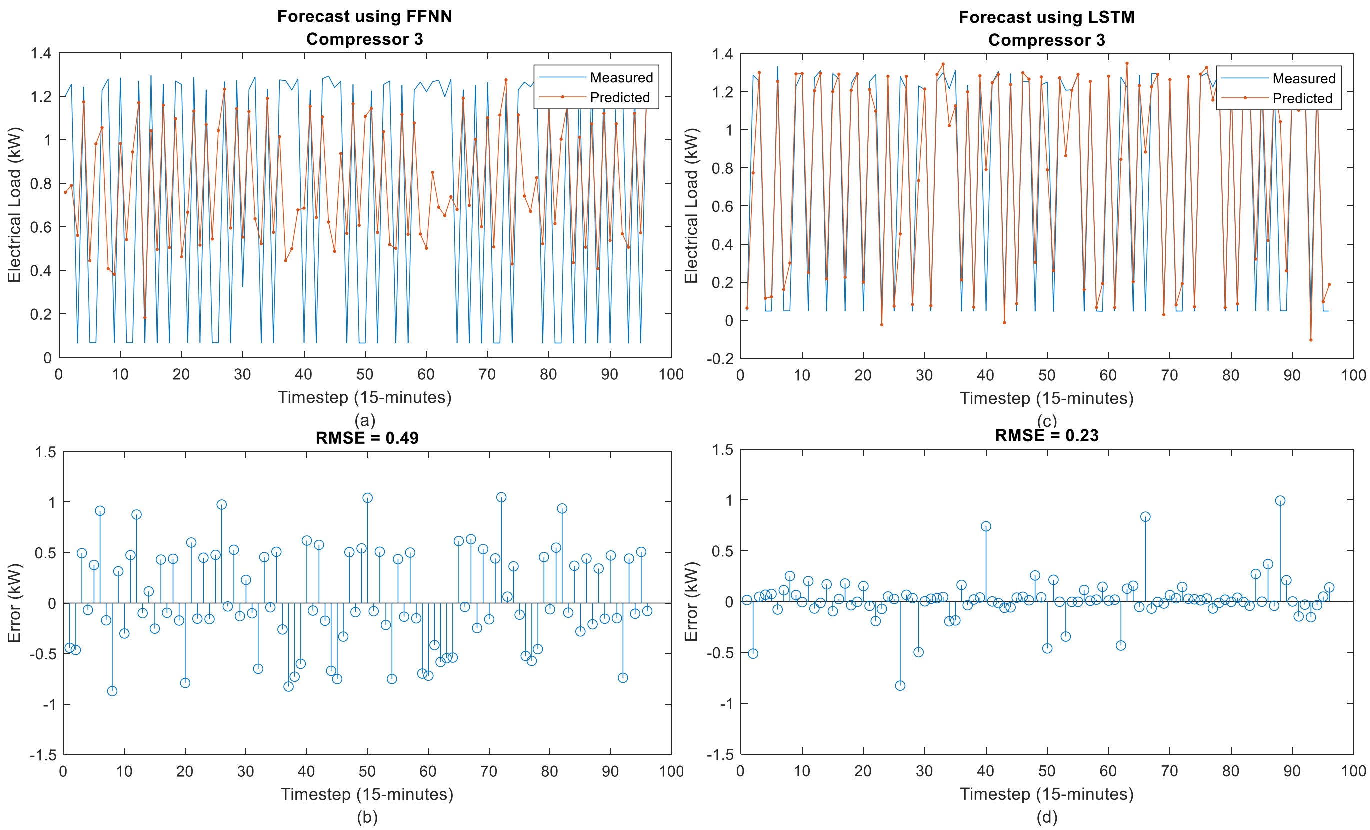

Fig. 12. Best prediction and their respective errors to the measurement in 15 minutes interval using FFNN and LSTM for Compressor 3.

Acknowledgements

This work was supported by the United States Department of Energy (grant number DE-EE0007711).

References

Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & Industrial Engineering*, *143*, 106435. https://doi.org/10.1016/j.cie.2020.106435

Atef, S., & Eltawil, A. B. (2020). Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electric Power Systems Research*, *187*, 106489. https://doi.org/10.1016/j.epsr.2020.106489

Aydinalp, M., Ismet Ugursal, V., & Fung, A. S. (2004). Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, *79*(2), 159–178. https://doi.org/10.1016/j.apenergy.2003.12.006

Bhaskar, K., & Singh, S. N. (2012). AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network. *IEEE Transactions on Sustainable Energy*, *3*(2), 306–315. https://doi.org/10.1109/TSTE.2011.2182215

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008). Time Series Analysis. In *Time Series Analysis: Forecasting and Control: Fourth Edition*. Wiley. https://doi.org/10.1002/9781118619193

Broomhead, D. S., & Lowe, D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems*, *2*, 321–355.

Brouwer, R. K. (2002). A feed-forward network for input that is both categorical and quantitative. *Neural Networks*, *15*(7), 881–890. https://doi.org/10.1016/S0893-6080(02)00090-4

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Dash, P. (1995). Building a Fuzzy Expert System for Electric Load Forecasting Using a Hybrid Neural Network. *Expert Systems with Applications*, *9*(3), 407–421. https://doi.org/10.1016/0957-4174(95)00013-Y

De Felice, M., Alessandri, A., & Catalano, F. (2015). Seasonal climate forecasts for medium-term electricity demand forecasting. *Applied Energy*, *137*, 435–444. https://doi.org/10.1016/j.apenergy.2014.10.030

Deb, C., Eang, L. S., Yang, J., & Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings*, *121*, 284–297. https://doi.org/10.1016/j.enbuild.2015.12.050

Fawzy, A., Mokhtar, H. M. O., & Hegazy, O. (2013). Outliers detection and classification in wireless sensor networks.

Egyptian Informatics Journal, *14*(2), 157–164. https://doi.org/10.1016/j.eij.2013.06.001

Fei, J., Zhao, N., Shi, Y., Feng, Y., & Wang, Z. (2016). Compressor performance prediction using a novel feed-forward neural network based on Gaussian kernel function. *Advances in Mechanical Engineering*, *8*(1), 168781401662839. https://doi.org/10.1177/1687814016628396

Firat, M., Turan, M. E., & Yurdusev, M. A. (2010). Comparative analysis of neural network techniques for predicting water consumption time series. *Journal of Hydrology*, *384*(1–2), 46–51. https://doi.org/10.1016/j.jhydrol.2010.01.005

Gers, F. A., Eck, D., & Schmidhuber, J. (2002). *Applying LSTM to Time Series Predictable Through Time-Window Approaches* (pp. 193–200). Springer. https://doi.org/10.1007/978-1-4471-0219-9_20

Gholamrezaei, M., & Ghorbanian, K. (2007). Rotated General Regression Neural Network. *2007 International Joint Conference on Neural Networks*, 1959–1964. https://doi.org/10.1109/IJCNN.2007.4371258

Ghorbanian, K., & Gholamrezaei, M. (2009). An artificial neural network approach to compressor performance prediction. *Applied Energy*, *86*(7–8), 1210–1221. https://doi.org/10.1016/j.apenergy.2008.06.006

Grolinger, K., L'Heureux, A., Capretz, M. A. M., & Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings*, *112*, 222–233. https://doi.org/10.1016/j.enbuild.2015.12.010

Hampel, F. R. (1974). The Influence Curve and Its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*(346), 383. https://doi.org/10.2307/2285666

Haykin, S. (2007). *Neural networks: a comprehensive foundation*. Prentice-Hall, Inc.

Hecht-Nielsen. (1989). Theory of the backpropagation neural network. *International Joint Conference on Neural Networks*, 593–605 vol.1. https://doi.org/10.1109/IJCNN.1989.118638

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*. https://doi.org/10.1109/59.910780

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, T. S., Lam, K. C., & Ng, S. T. (2001). River flow time series prediction with a range-dependent neural network. *Hydrological Sciences Journal*, *46*(5), 729–745. https://doi.org/10.1080/02626660109492867

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jovanović, R. Ž., Sretenović, A. A., & Živković, B. D. (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, *94*, 189–199. https://doi.org/10.1016/j.enbuild.2015.02.052

Kalogirou, S. A., & Bojic, M. (2000). Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*. https://doi.org/10.1016/S0360-5442(99)00086-9

Kamel, N., & Baharudin, Z. (2007). Short term load forecast using Burg autoregressive technique. *2007 International Conference on Intelligent and Advanced Systems*, 912–916. https://doi.org/10.1109/ICIAS.2007.4658519

Kandil, M. S., El-Debeiky, S. M., & Hasanien, N. E. (2002). Long-term load forecasting for fast developing utility using a knowledge-based expert system. *IEEE Transactions on Power Systems*, *17*(2), 491–496. https://doi.org/10.1109/TPWRS.2002.1007923

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, *4*, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968

Kim, P. (2017). MATLAB Deep Learning. In *MATLAB Deep Learning*. Apress. https://doi.org/10.1007/978-1-4842-2845-6

Ledesma, S., Belman-Flores, J. M., & Barroso-Maldonado, J. M. (2015). Analysis and modeling of a variable speed reciprocating compressor using ANN. *International Journal of Refrigeration*, *59*, 190–197. https://doi.org/10.1016/j.ijrefrig.2015.08.009

Lee, C.-M., & Ko, C.-N. (2011). Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications*, *38*(5), 5902–5911. https://doi.org/10.1016/j.eswa.2010.11.033

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*(2), 164–168. https://doi.org/10.1090/qam/10666

Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & Chemical Engineering*, *28*(9), 1635–1647. https://doi.org/10.1016/j.compchemeng.2004.01.009

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

Mason, K., Duggan, J., & Howley, E. (2018). Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks. *Energy*, *155*, 705–720. https://doi.org/10.1016/j.energy.2018.04.192

Memarzadeh, G., & Keynia, F. (2020). A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets. *Energy Conversion and Management*, *213*, 112824. https://doi.org/10.1016/j.enconman.2020.112824

Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings*, *40*(12), 2169–2176.

https://doi.org/10.1016/j.enbuild.2008.06.013
Niu, D., Wang, Y., & Wu, D. D. (2010). Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2009.08.019
Nolde, K., & Morari, M. (2010). Electrical load tracking scheduling of a steel plant. *Computers & Chemical Engineering*, *34*(11), 1899–1903. https://doi.org/10.1016/j.compchemeng.2010.01.011
Pearson, R. K., Neuvo, Y., Astola, J., & Gabbouj, M. (2016). Generalized Hampel Filters. *EURASIP Journal on Advances in Signal Processing*, *2016*(1), 87. https://doi.org/10.1186/s13634-016-0383-6
Platon, R., Dehkordi, V. R., & Martel, J. (2015). Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. *Energy and Buildings*, *92*, 10–18. https://doi.org/10.1016/j.enbuild.2015.01.047
Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, *50*, 1352–1372. https://doi.org/10.1016/j.rser.2015.04.065
Ribeiro, G. T., Mariani, V. C., & Coelho, L. dos S. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Engineering Applications of Artificial Intelligence*, *82*, 272–281. https://doi.org/10.1016/j.engappai.2019.03.012
Rodrigues Moreno, S., Gomes da Silva, R., Cocco Mariani, V., & dos Santos Coelho, L. (2020). Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. *Energy Conversion and Management*, *213*, 112869. https://doi.org/10.1016/j.enconman.2020.112869
Saidur, R., Rahim, N. A., & Hasanuzzaman, M. (2010). A review on compressed-air energy use and energy savings. *Renewable and Sustainable Energy Reviews*, *14*(4), 1135–1153. https://doi.org/10.1016/j.rser.2009.11.013
Schulze, M., Nehler, H., Ottosson, M., & Thollander, P. (2016). Energy management in industry - A systematic review of previous findings and an integrative conceptual framework. *Journal of Cleaner Production*, *112*, 3692–3708. https://doi.org/10.1016/j.jclepro.2015.06.060
Shekhar, S., & Amin, M. B. (1992). Generalization by neural networks. *IEEE Transactions on Knowledge and Data Engineering*, *4*(2), 177–185. https://doi.org/10.1109/69.134256
Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., & Woo, W. (2015). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Nips*, 802–810. http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting
Specht, D. F. (1991). A General Regression Neural Network. *IEEE Transactions on Neural Networks*. https://doi.org/10.1109/72.97934
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.
Tian, Z., Gu, B., Yang, L., & Lu, Y. (2015). Hybrid ANN–PLS approach to scroll compressor thermodynamic performance prediction. *Applied Thermal Engineering*, *77*, 113–120. https://doi.org/10.1016/j.applthermaleng.2014.12.023
Vu, D. H., Muttaqi, K. M., & Agalgaonkar, A. P. (2015). A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. *Applied Energy*, *140*, 385–394. https://doi.org/10.1016/j.apenergy.2014.12.011
Weninger, F., Erdogan, H., Watanabe, S., Vincent, E., Le Roux, J., Hershey, J. R., & Schuller, B. (2015). Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR. *Latent Variable Analysis and Signal Separation*, 91–99. https://doi.org/10.1007/978-3-319-22482-4_11
Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560. https://doi.org/10.1109/5.58337
Wong, S. L., Wan, K. K. W., & Lam, T. N. T. (2010). Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, *87*(2), 551–557. https://doi.org/10.1016/j.apenergy.2009.06.028
Wu, D. C., Amini, A., Razban, A., & Chen, J. (2018). ARC algorithm: A novel approach to forecast and manage daily electrical maximum demand. *Energy*, *154*, 383–389. https://doi.org/10.1016/j.energy.2018.04.117
Wu, Q. (2010). Power load forecasts based on hybrid PSO with Gaussian and adaptive mutation and Wv-SVM. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2009.05.011
Yona, A., Senjyu, T., Saber, A. Y., Funabashi, T., Sekine, H., & Kim, C.-H. (2007). Application of Neural Network to One-Day-Ahead 24 hours Generating Power Forecasting for Photovoltaic System. *2007 International Conference on Intelligent Systems Applications to Power Systems*, 1–6. https://doi.org/10.1109/ISAP.2007.4441657
Yu, Y., Chen, L., Sun, F., & Wu, C. (2007). Neural-network based analysis and prediction of a compressor's characteristic performance map. *Applied Energy*. https://doi.org/10.1016/j.apenergy.2006.04.005
Zadeh, L. A. (1988). Fuzzy logic. *Computer*, *21*(4), 83–93. https://doi.org/10.1109/2.53
Zahlan, J., & Asfour, S. (2015). A multi-objective approach for determining optimal air compressor location in a manufacturing facility. *Journal of Manufacturing Systems*, *35*, 176–190. https://doi.org/10.1016/j.jmsy.2015.01.003