Vol. 1, No. 2, 2023
ISSN: 2987-3932
https://pubs.ast-ptm.or.id/index.php/eat

Load Forecasting on 150 kV Grid Substation Based on Spatiotemporal Deep Neural Network

Karisma Trinanda Putra [a,1,*], Duta Fahri Alfiansyah [b,2], Muhtadan [a,3], Sunardi [a,4], Ramadoni Syahputra [c,5]

[a] Electrical Engineering Department, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia
[b] Department of Electronics and Instrumentation Indonesian Nuclear Technology Polytechnic Yogyakarta, Indonesia
[c] Mechanical Engineering Department, Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia
[1] karisma@ft.umy.ac.id *; [2] duta.fahri.ft19@mail.umy.ac.id; [3] muhtadan@brin.go.id; [4] sunardi@mail.umy.ac.id;
[5] ramadoni@umy.ac.id
* Corresponding Author

ABSTRACT

A smart grid concept with a prediction system provides accurate information as an early warning in the process of generation and distribution of electrical energy. The complexity of the power distribution system that involves complex parameter settings is a major challenge that is difficult to predict. Although the data recording in the smart grid system is done centrally, however, the prediction system faces a lack of accuracy due to incomplete records and low data quantity. In this paper, a spatiotemporal deep neural network is developed based on convolutional long short-term memory (Conv-LSTM) to extract long-term short-term patterns in an electrical load dataset. This data set is obtained from daily measurements for six months at Cawang Baru Substation, Indonesia. The proposed model adopts the basic concept of multi-layer perceptron to record temporal patterns in several stages, thereby producing more accurate results. This model uses supervised learning techniques to propagate sequential data into the target which is the next event of the data series. Furthermore, the proposed architecture supports multivariate feature extraction so as to capture important correlations between multi-dimensional features. This study also uses the basic Multivariate LSTM (MV-LSTM) model and naive Machine Learning models including Logistic Regression (LR), Random Forest (RF), and Support Vector Regression (SVR) as benchmarking methods for the proposed model. In the testing process, Conv-LSTM achieves higher accuracy than MVLSTM, SVR, LR, and RF with scores of 0.3688, 0.3645, 0.1332, 0.1438, and 0.1234, respectively, evaluated using R-squared. Finally, experimental results support the view that combining multivariate data and a spatiotemporal prediction model is superior for time series prediction tasks rather than univariate data.

KEYWORDS
Spatiotemporal deep neural
Electrical load dataset
Time series prediction

1. Introduction

In recent decades, with rapid population growth, industrialization, and traffic, electricity has become a major need in modern cities, especially in Asia. A report by an electricity company in Indonesia estimates that electricity consumption will increase by an average of 8.42% per year with more than 79 million people having electricity by 2020 [1]. Meanwhile, a statistical revealed that the increase in SAIFI almost doubled between 2014-2018 [2]. The growth in electricity consumption must also be accompanied by an increase in the reliability of electricity distribution because the industrial sector is estimated to suffer losses of up to 245 billion rupiah. Therefore, to minimize the effects of blackouts, an electrical energy forecasting system needs to be developed by involving several regulatory parameters that greatly affect the reliability of the electricity distribution process.

To reduce uncontrolled blackouts, several countries have developed power generation sites, distribution systems, and electrical energy transmission systems that are interconnected in a smart grid. Currently, smart grid technology is developing with a data monitoring system that records daily electrical load conditions [3]. Furthermore, this automatic system generates large amounts of data and includes several critical parameters in the process of distributing electrical energy. This electrical data can potentially be used as a valuable source for predicting future electricity conditions. Data from monitoring sites can be accessed and collaborated with other data from other monitoring sites. In general, this massive scale data can provide better global prediction results as early information for electricity providers about the condition of electricity distribution [4].

Advances in machine learning (ML) algorithms may be the solution to the aforementioned problems. These algorithms have evolved providing a better prediction performance especially in extracting sequential patterns. In addition, this dataset contains patterns in the spatial domain that describe the correlation between adjacent measurements. By using ML models, it is possible to extract long short-term temporal patterns with greater precision. Theoretically, a prediction model based on modern ML models (i.e., neural network-based) would be more precise because the latest advances provide a complex multi-dimensional feature extraction [5]. Long-term dependencies in the input sequences can be tracked using Long Short-Term Memory (LSTM) and then propagate onto the target, which is the next event in the data series. These advances, in conjunction with the massive multi-dimensional dataset, could improve the overall performance of the prediction system.

This study proposes an ML model based on Convolutional Long Short-Term Memory (Conv-LSTM) to recognize short and long-term temporal patterns from the real-world electrical load dataset collected from the real-world experiments in Cawang Baru Substation, Indonesia. This dataset inherits the fundamental concepts of smart grid, which covers measurements providing monitoring capability in the Substation. By using multi-layer perceptron, the patterns are extracted using several stages achieving a consistent training mechanism for each layer. Combining LSTM in several layers allows the proposed model to extract numerous and complex pre-trained patterns [6]. First, the dataset is divided into two subsets i.e., training set and testing set. Then, the training set is trained in the model using back-propagation mechanism. These processes of updating knowledge (i.e., in the form of parameter weights) are iterated in several back-propagation mechanisms converging the outputs to match the targets. It means that the prediction results equal to the predefined target for all data set. Moreover, the proposed model can be utilized for developing smart grid, especially in Indonesia. The government could access the datasets and prediction results in real-time, raising authorities' awareness of the next events related to electrical distribution systems. The ML-based prediction systems may be the solution to the implementation of the implementation of smart grid. In addition, by using modern ML models, long short-term temporal patterns could be extracted in detail. The accuracy of a prediction model based on the combination of multivariate dataset and LSTM is improved because each data point is gradually propagated using several stages of LSTM layers. MLP can propagate the input onto the target using backpropagation algorithm while LSTM can extract long short-term dependencies in the data series.

The rest of this paper is outlined as follows. Literature studies are described in Section II, followed by the ML methods in Section III. The next section IV describes the experimental results. In the end, a brief conclusion is summarized in Section V.

2. Previous Works

Forecasting of electrical load is important to do to anticipate unexpected conditions so as to increase the awareness of the electricity distribution operator. Variations in parameters in the distribution process, such as active power, reactive power, and load current, can be used as indicators to forecast the active power for several periods in the future. This variation is characterized by a high proportion of outliers and constant changes from time to time of the parameters. These parameters have a strong relationship so that it is suitable to be used as an input for a multivariate prediction system. Currently, there are few models [7] that can be used to predict multivariate data, namely models based on artificial neural networks. The larger the number of datasets collected, the more varied the patterns that can be captured by the ANN model so as to increase the prediction accuracy. Some studies predict data sequences using clustered data to simplify the process and better record temporal patterns. By using clustering technique, the data is reorganized into smaller domains. The clustering method is particularly useful for large-scale predictive models, albeit at the expense of resolution. By using linear regression analysis, researchers can make forecasting models with a limited number of clusters. However, this method cannot capture more complicated characteristics [8]. The longer the time period to be predicted, the greater the difference between the prediction and the actual data.

Some researchers use perceptrons to analyze signal complexity with a high proportion of outliers. However, the concept is only implemented on a small one-dimensional dataset, without involving long-term data acquisition. However, this method can only be applied to data with small series and moderate complexity. The complexity of the data set features makes it more difficult for the perceptron to achieve convergent learning. The discovery of the LSTM network improves the performance of the prediction system for forecasting time series data with high complexity [9]. It has been shown that LSTM improves the performance of time series data prediction systems. The temporal pattern contains a strong correlation of data between the current measurement and the previous measurement. Using the LSTM, these temporal changes can be extracted precisely. Previous studies have discussed the ability of the LSTM model to predict sequence data. There is potential for this learning model to be used to analyze the temporal relationship between long-term measurements of multivariate variables. The proposed ConvLSTM model can extract long-term short-term patterns in the electrical load data set by using three features, namely active power, reactive power, and load current. In addition, ConvLSTM is capable of processing multiple input channels, enabling its ability to extract multiple features in multivariate data sets better than univariate models.

3. Methodology

For evaluating the performance of the proposed model, three naive ML approaches are described as follows. The first subsection provides descriptions of Logistic Regression model. In the next subsection, a Random Forrest model is presented. In the end, Support Vector Machine is briefly described.

3.1. Logistic Regression

Logistic Regression (LR) does not require a linear relationship between the independent variable and the dependent variable. The independent variable does not require the assumption of multivariate normality (assumptions are normal). The assumption of homoscedasticity is not required. The dependent variable must be dichotomous (2 categories, for example: high and low or good and bad). Independent variables do not have to have the same diversity between groups of variables. The categories in the independent variables must be separate from each other or are exclusive. The sample required is relatively large, a minimum of up to 50 data samples is required for a predictor variable (independent). The logistic regression method can select the relationship because it uses a non-linear log transformation approach to predict the odds ratio. Odd ratios in logistic regression are often expressed as probabilities.

$$\pi(x) = \frac{exp(\beta_0+\beta_1 x)}{1+exp(\beta_0+\beta_1 x)} \tag{1}$$

where, π, B0, B1, x, are, probability of occurrence, constant, regression coefficient, and independent variable, respectively.

3.2. Random Forest

Random Forest (RF) is a supervised learning algorithm released in the early 2000s. Random Forests are commonly used to solve problems related to classification, regression, and so on. There are two things that make this algorithm called random, namely: (1) Each tree grows on a different bootstrap sample taken from the training data randomly. (2) In each split node during decision tree formation, a sample portion of m variables is selected from the original data set and then the best one will be used in that node. This algorithm is a combination of several tree predictors or can be called decision trees where each tree depends on the random vector value which is sampled freely and evenly on all trees in the forest. Prediction results from Random Forest are obtained through the highest results from each individual decision tree (voting for classification and average for regression). For RF consisting of N trees, it is formulated as stated in Equation (2).

$$l(y) = argmax_c(\sum_{n=1}^{N} I_{h_n(y)=c}) \tag{2}$$

where, I is the indicator function and is the n-th tree of RF.

3.3. Support Vector Regression

Both classification and regression tasks can be solved using support vector machine (SVM). This study focuses on SVM for regression tasks i.e., known as support vector regression (SVR). The model produced by SVR relies only on a subset of the training data, as the cost function disregards samples whose prediction is close to the target. A linear SVR works by making a straight line (i.e., known as hyperplane) between two classes. That means all the data points on one side of the line will represent a category and the data points on the other side of the hyperplane will represent another category. There can be an infinite number of hyperplanes to choose from, but the optimal one can be achieved using iterated optimizing procedures. Linear SVR algorithm is also one type of ML algorithm which performs better because it chooses the optimal hyperplane to classify data points. Equation (2) estimates a hyperplane that separates the data points and is as optimal as possible from the closest data points.

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (3)$$

where, $a_i, a_i *$ are the Lagrange multipliers, $k = x_i, x$ denotes kernel function, and b is the bias.

3.4. Long Short-Term Memory Network

Long-Short Term Memory (LSTM) has been successfully used in sequential data prediction tasks. To control the value of the cell state c LSTM has two gates as the inputs as shown in Figure 1. Forget gate f determines how much of the cell state at the previous time $c_t - i$ is retained to the current cell state c_t. Meanwhile, input gate controls how much input at the current time is saved to the current unit state. In each process of updating the LSTM cell, the cell accepts the hidden state of the previous cell and the input. Then, the cell determines whether or not to retain the specific calculation process of its inputs. The forget gate calculates the previous state using Equation (4), where is the sigmoid activation function and is the weight of forget gate. denotes the vectors which are connected into a longer vector while determines the bias term of the forget gate. How much information in the current cell input needs to be saved to the current state of the LSTM cells, is calculated using Equation (5). The sigmoid layer of the input gate layer process how information is updated to the output linearity. Finally, Equation (6) is used to update the cell which can be considered as the current output using activation function.

$$f_t = \zeta\left(W_f \cdot \left(h_{t-1}, x_t\right) + b_f\right) \quad (4)$$

$$i_t = \sigma\left(W_i \cdot \left(h_{t-1}, x_t\right) + b_i\right) \quad (5)$$

$$\tilde{c}_t = tanh\left(W_c \cdot \left(h_{t-1}, x_t\right) + b_c\right) \quad (6)$$

3.5. The Proposed Multivariate LSTM Network

Multivariate data analysis is research that started with application of new computational techniques that allow analysis of more than one dimension. LSTM network supports multi-channel input. This means, several features in sequential data can be processed simultaneously and will affect the propagation of input to output. Multivariate data has several features with potentially valuable temporal patterns for generating knowledge convergence in LSTM networks. Therefore, it is important to record multivariate data and theoretically produce better performance than univariate data.

Data measurements was recorded at PT. PLN GIS Cawang Baru, East Jakarta City, Special Capital Region of Jakarta. From these records, several data were obtained including data on daily active loads, reactive loads, and load currents on a 500/150 kV power transformer unit for 6 months, from April to September 2020. The data that has been obtained will be used to observe conditions. electrical load connected to the power transformer and estimate the electrical load. The dataset is then collected and further processed as shown in Figure 2.

First, preprocessing is performed to produce continuous data and reduce missing information by averaging technique. Then, the data is scaled between the range 0 to 1 according to the input capacity of the LSTM network. After that, the data is organized into sub-sets of historical data. The data subset consists of input and output parts. The input part is daily recording data for the past 3 weeks (xt-1, xt-2, …, xt-27) while the output part is the current day data (xt). This subset of data is divided into 2 processes for training and testing procedures with a ratio of 80:20, respectively. In the training process, the data is trained using 5 types of ML models, namely LR, RF, SVR, MV-LSTM, and Conv-LSTM. At the end, each model is tested using data testing so as to produce performance matrices, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2).

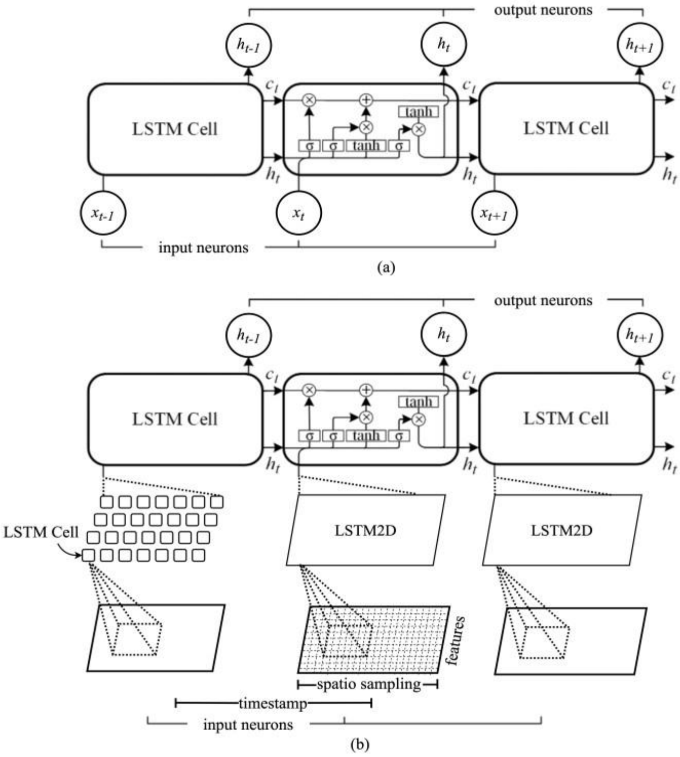

Fig. 1. Long Short-Term Memory network utilizes LSTM cels which are sequentially arranged to extract temporal feature inside a spatiotemporal dataset. (a) Multivariate LSTM (MV-LSTM). (b) Convolutional LSTM (Conv-LSTM).

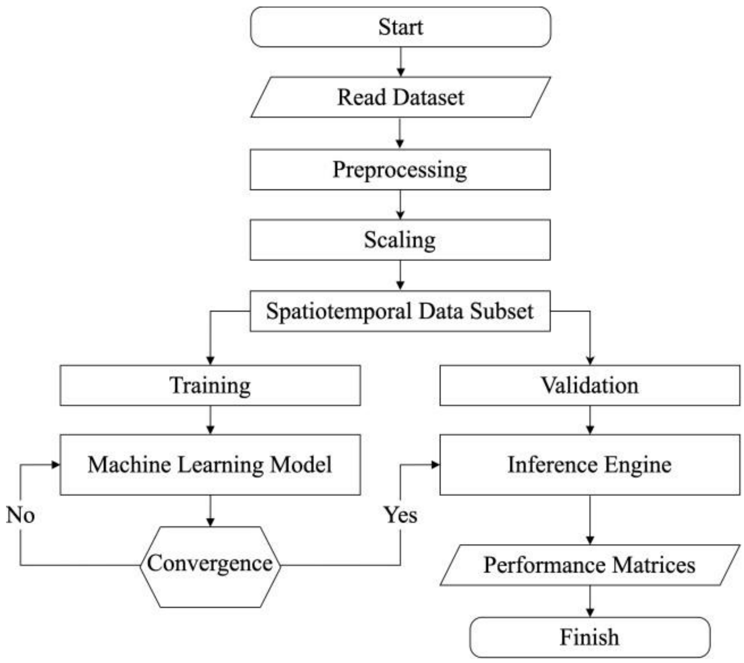

Fig. 2. Diagram block of the training procedures which involve ML models. The testing procedures is performed using three performance matrices including MAE, RMSE, and R^2.

4. Result

The performance matrices are used to evaluate the performance of the training process including MAE, RMSE, and R2, formulated in Equation (7), (8), and (9), respectively.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|o_{target(i)} - o_{predicted(i)}\right| \tag{7}$$

$$RMSE = \sqrt{\sum_{t=1}^{T}\frac{\left(o_{target(i)} - o_{predicted(i)}\right)^2}{T}} \tag{8}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{9}$$

where t arg et i( ) o is the ground truth data, predicted i( ) o is the predicted load, and i is index timestamp of the data, RSS is sum of squares of residuals, and TSS is total sum of squares.

Experiments are carried out by testing the model that has been trained using subset testing. In general, the model training process on the LR, RF, and SVR models is faster to complete because it only involves variables in mathematical equations. In contrast, in the LSTM model models, the intelligence function using a large number of neurons so that a total of 71,051 parameters are included. Furthermore, each of these parameters (i.e., model weight) is trained repeatedly until it reaches the specified epoch. This back propagation process takes longer than the traditional ML approach. However, the inference process in all ML models has almost the same execution time because it only involves the forward propagation calculation process. With subset testing, predictions can be generated to test how reliable the system is when faced with a new dataset that has not been trained before as qualitatively shown in Figure 3 to 6. The testing procedure produces data with different accuracy for each model. The naive RF model produces the worst performance of the five models tested although it can still extract sequence patterns in the test subset set. Furthermore, the LR model is visually comparable to the predicted results propagated by the SVR model. The RF model is slightly more consistent in producing stable predictions. The RF models has more capability to accommodate temporal feature extraction in the dataset. Although only one feature is used as input for the prediction model in naive ML models, these temporal patterns are fundamentally able to produce accurate, consistent, and reliable predictions. In smart grid concept, the data recording process includes many features and in this massive data, valuable knowledge is stored. This is in line with the physical properties of electrical load which are easily influenced by changes in other features. Therefore, a prediction system that is able to accommodate multivariate datasets is certainly more feasible to implement.

MAE, RMSE, and R2 evaluate the prediction error with the actual measurement. RMSE is more widely used than MAE to evaluate the performance of the regression model. RMSE is used as the default metric to calculate the loss function in each iteration in MLP although, it is more complex than MAE. In general, a lower RMSE values imply better regression model accuracy. However, a higher R2 value in the prediction model is considered more feasible. R2 is used to explain how well the independent variables in the regression model explain the variability in the dependent variable. The value of R squared always increases with the addition of independent variables which may lead to the addition of redundant variables in the prediction model. As shown in Table 1, the RF model produces the lowest performance as evidenced by the lower MAE and RMSE compared to the other models. Even though LR and SVR have higher MAE and RMSE scores than RF, the R2 score of this model is higher than that produced by RF. Thus, RF is not suitable to be used for sequential feature extraction. On the other hand, the LSTM-based ML models have R 2 scores higher than the naive ML models. Another advantage is that LSTM-based ML model can be organized into deeper layers to accommodate the complexity of the variables being trained. LSTM also supports multivariate features so that it is more feasible to be implemented for developing smart grid concept. Based on experiment on Table 1, in the training process, Conv-LSTM produces higher accuracy than Conv-LSTM, SVR, LR, and RF with R2 score of 0.9936, 0.9874, 0.1811, 0.2164, 0.6493, respectively. In the testing process, Conv-LSTM achieves higher accuracy than MV-LSTM, SVR, LR, and RF with R2 score of 0.3688, 0.3645, 0.1332, 0.1438, 0.1234, respectively.

Table 1. Specification of the Experimental Models

Phase	Metrics	Conv-LSTM	MV-LSTM	SVR	LR	RF
Training	MAE	0.1757	0.2574	2.0716	2.0252	**1.3763**
	RMSE	0.2372	0.3319	2.6792	2.6208	**1.7533**
	R^2	**0.9936**	0.9874	0.1811	0.2164	0.6493
Testing	MAE	2.8472	3.1651	**2.4950**	2.6446	2.5758
	RMSE	3.8966	3.9487	3.5984	3.6153	**3.5829**
	R^2	**0.3688**	0.3645	0.1332	0.1438	0.1234

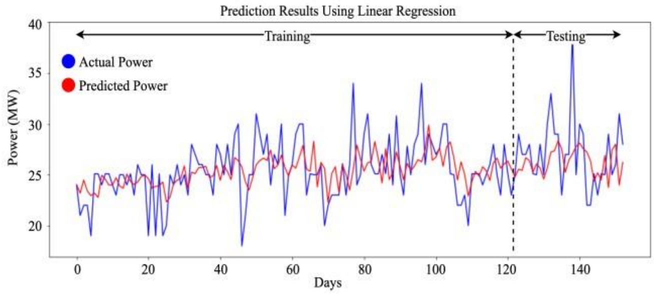

Fig. 3. Diagnostic line plot showing training and testing evaluation of the LR model.

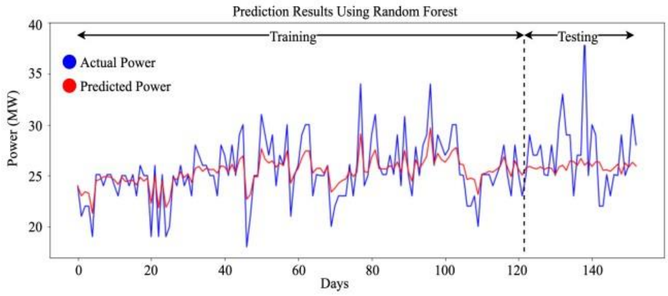

Fig. 4. Diagnostic line plot showing training and testing evaluation of the RF model.

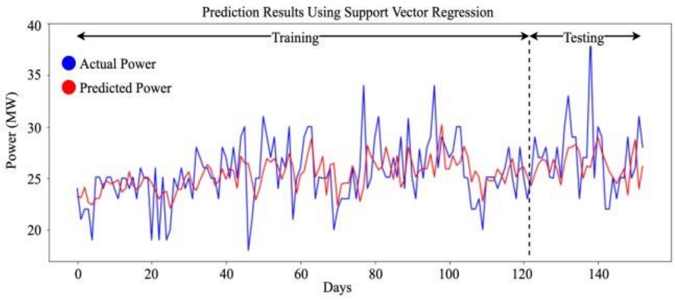

Fig. 5. Diagnostic line plot showing training and testing evaluation of the SVR model.

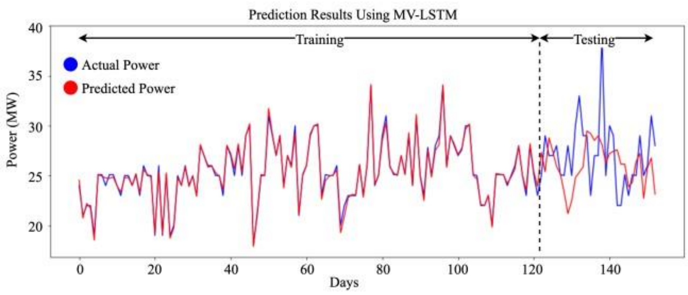

Fig. 6. Diagnostic line plot showing training and testing evaluation of the MV-LSTM model.

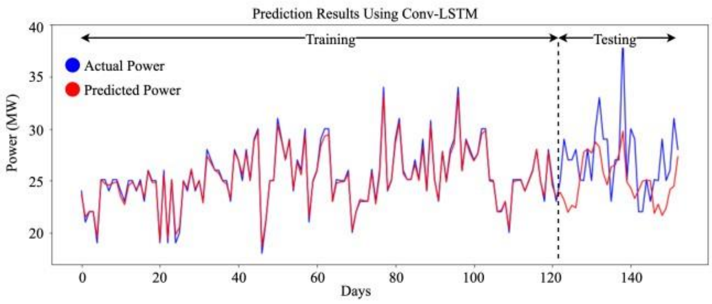

Fig. 7. Diagnostic line plot showing training and testing evaluation of the Conv-LSTM model.

5. Conclusion

To reduce the risk of unexpected events in electrical energy distribution processes, the discovery and prediction of electrical load patterns is needed. In the literature on finding and predicting electrical load patterns through machine learning, previous research only focused on finding temporal patterns without considering changes in other parameters that might be used as clues to improve model performance. This paper complements the shortcomings of previous research, which established the complete structure of the discovery and prediction of next-day electrical loads using a multivariate LSTM architecture. Through this architecture, the proposed scheme can analyze and predict the next day's electrical load using a multivariate temporal periodic pattern on a daily data record at the Cawang Baru Substation. The results showed that in the testing procedure, the proposed Conv-LSTM network was superior in producing better prediction performance than the MV-LSTM, SVR, LR, and RF with R 2 score of 0.3688, 0.3645, 0.1332, 0.1438, 0.1234, respectively. In addition, there are promising studies on the findings of this multivariate pattern. This possibility is to find a more suitable and effective pattern extraction to improve the performance of the proposed structure such as by using spatiotemporal feature extraction.

Acknowledment

This work was supported by Universitas Muhammadiyah Yogyakarta through Center of Artificial Intelligence and Robotic Studies, Research and Innovation Institute of UMY under the grant numbers 01/RIS-LRI/I/2022.

References

[1] Konsumsi Listrik per Kapita (MWH/Kapita), available online: https://www.bps.go.id/indicator/7/1156/1/konsumsi-listrik-perkapita.html (accessed on August 9, 2022).

[2] Statistik PLN 2018, available online: https://web.pln.co.id/statics/uploads/2019/07/Buku-Statistik-2018-vsIndo-Final.pdf (accessed on August 9, 2022).

[3] T. S. Bomfim, "Evolution of Machine Learning in Smart Grids," 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), 2020, pp. 82-87.

[4] L. Fan, J. Li, Y. Pan, S. Wang, C. Yan and D. Yao, "Research and Application of Smart Grid Early Warning Decision Platform Based on Big Data Analysis," 2019 4th International Conference on Intelligent Green Building and Smart Grid, 2019, pp. 645-648.

[5] F. D. Syahfitra, R. Syahputra, and K. T. Putra, "Implementation of Backpropagation Artificial Neural Network as a Forecasting System of Power Transformer Peak Load at Bumiayu Substation," Journal of Electrical Technology UMY, vol. 1, no. 3. Universitas Muhammadiyah Yogyakarta, 2017.

[6] R. Zhichao, C. Chao, D. Yingying, Z. Wentao, W. Jun and Z. Ruixiao, "Short-term load forecasting of multi-layer LSTM neural network considering temperature fuzzification," 2020 IEEE Sustainable Power and Energy Conference, 2020, pp. 2398-2404.

[7] R. Chuentawat and Y. Kan-ngan, "The Comparison of PM2.5 forecasting methods in the form of multivariate and univariate time series based on Support Vector Machine and Genetic Algorithm," 2018 15th International Conference

on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON), 2018, pp. 572-575.

[8] S. Kumar, S. Mishra and S. Gupta, "Short Term Load Forecasting Using ANN and Multiple Linear Regression," 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), 2016, pp. 184-186.

[9] K. T. Putra, H. -C. Chen, Prayitno, M. R. Ogiela, C. -L. Chou, Weng, C. -E. Weng, Z. -Y. Shae, “Federated Compressed Learning Edge Computing Framework with Ensuring Data Privacy for PM2.5 Prediction in Smart City Sensing Applications,” Sensors, vol. 21, no. 13. MDPI AG, p. 4586, Jul. 04, 2021.