Predicting Electrical Consumption of Seven Countries Using Neural Network and Linear Regression

[1]Dina Mariani, [2]Yanuar Mahfudz Safarudin

[1] Electrical Engineering Department,
Muhammadiyah University of Semarang, Indonesia
[2] Electrical and Communication Engineering Department,
College of Engineering, United Arab Emirates University, United Arab Emirates
[2] Energy Generation Engineering Technology,
Department of Mechanical Engineering, Politeknik Negeri Semarang, Indonesia

[1]dinamariani@unimus.ac.id, [2]700043219@uaeu.ac.ae

***Abstract*— This project proposes the forecast of electrical consumption from seven countries around the world. The data are open accessed from The International Energy Agency (https://www.iea.org/). Overall, there are 791 data consisting seven countries from January 2014 to May 2023. This project presents a proposal to utilize a neural network as the primary model for predicting electrical consumption, conducted. with a secondary model of linear regression employed for the purpose of comparison**

***Keywords*— Artificial neural network,forecast,hidden layers, load consumption**

I. INTRODUCTION

This project proposes the forecast of electrical consumption from seven countries around the world. The data are open accessed from The International Energy Agency (https://www.iea.org/) [1][2]. Overall, there are 791 data consisting seven countries from January 2014 to May 2023 [3]. Every nation exhibits a unique and distinct pattern of electrical consumption [4]. For instance, countries located in the northern hemisphere, such as Canada and Norway, experience winter during the month of December [5]. Conversely, countries located in the southern hemisphere, like Australia and Chile, undergo winter in July. Tropical countries namely Colombia, on the other hand, experience no winter [6]. Given the analysis of trends ranging from 2014 to 2023, the model should possess the capability to predict the electrical consumption in those mentioned seven nations for the year 2024. To solve this challenge, neural networks [7]-[9] and linear regression [10]-[14], that has been widely used to predict the electrical consumption [15][16], will be used to forecast the given problem.

II. METHOD

This project presents a proposal to utilize a neural network as the primary model for predicting electrical consumption, conducted. with a secondary model of linear regression employed for the purpose of comparison.

A. Architecture of Neural Network and Linear Regression

This project presents a proposal to utilize a neural network as the primary model for predicting electrical consumption, with a secondary model of linear regression employed for the purpose of comparison.

1) Neural Network Model: For the neural network model, there will be three inputs consisting the year, month, and country categories. The number of hidden layers is selected as two layers with seven neurons in every hidden layer, as shown in Figure 1. And then, the output will be the electrical consumption.

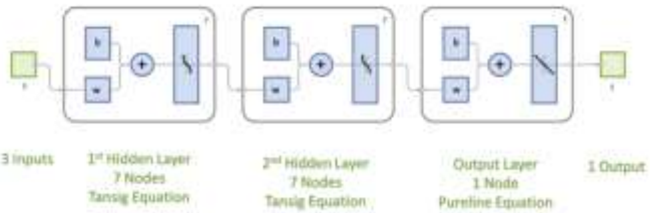

Figure 1 Illustration of Neural Network Design

As neural network consists of two hidden layers, and each of them will have seven neurons (nodes), there will be 7x3 (21) weights from the input to the hidden layer 1, 7x7 (49) weights from hidden layer 1 to hidden layer 2, and 1x7 weights from hidden layer 2 to the output with the total weight of 77. For the biases, there will be 15 biases (7 in 1st hidden layer, 7 in 2nd hidden layer, and 1 in the output). For the activation function, both of the hidden layers will use tansig function, whereas the output node will use pureline function. The selection of the number of hidden layers and nodes (7x7) for the neural network in this forecasting application was determined through a process of trial and error, resulting in the configuration that showed the lowest error rate.

2) Linear Regression Model: The results of the neural network will be evaluated with those of the linear regression method. In order to address the complexity of the regression problem, it is crucial to improve the input matrix beyond the incorporation of only three variables, specifically month, year, and countries. It is advisable to also incorporate the quadratic, cubic, and fourth power terms for each of these input variables. In addition, the additional input consists only 1 should be added for calculating the θo.

There will be several additional inputs corresponding the non-linear variables, and after several tries, 9 inputs will be chosen as:

1. Constant number of 1 (X_1)
2. Number from 0 to 1 representing month (X_2)
3. Number from 0 to 1 representing year (X_3)
4. Number from 0 to 1 representing countries (X_4)
5. Square number of month number representation ($X_5=X_2^2$)
6. Square number of year number representation ($X_6=X_3^2$)
7. Square number of country number representation ($X_7=X_4^2$)
8. Number representing the country to the power of 3 ($X_8=X_4^3$)
9. Number representing the country to the power of 4 ($X_9=X_4^4$)

The gradient of linear regression will be represented as θ_o to θ_8. Hence the output of y (electrical consumption) can be calculated using following regression equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{1.1} & x_{1.2} & x_{1.3} & x_{1.4} & x_{1.5} & x_{1.6} & x_{1.7} & x_{1.8} & x_{1.9} \\ x_{2.1} & x_{2.2} & x_{2.3} & x_{2.4} & x_{2.5} & x_{2.6} & x_{2.7} & x_{2.8} & x_{2.9} \\ x_{3.1} & x_{3.2} & x_{3.3} & x_{3.4} & x_{3.5} & x_{3.6} & x_{3.7} & x_{3.8} & x_{3.9} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n.1} & x_{n.2} & x_{n.3} & x_{n.4} & x_{n.5} & x_{n.6} & x_{n.7} & x_{n.8} & x_{n.9} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \end{bmatrix}$$

In training phase, the dataset consists of 672 rows. Therefore, the dimensions of the output matrix (y) will be 672x1, the dimensions of the input matrix (x) will be 672x9, and the dimensions of the gradient matrix will be 9x1. In the testing condition, a total of 119 data points will be used. Consequently, the resulting matrix will have dimensions of 119x1 for the output, 119x9 for the input, and the gradient matrix will retain its dimensions of 9x1.

B. Model Training

1) Dataset Detail: Data preprocessing will be conducted for both the input and output variables. Due to the natural limitations of neural networks and linear regression models, which are designed to process numerical data rather than textual representations such as months or country names (e.g., "January"), it is necessary to convert the month and country variables into numerical values ranging from 0 to 1. As an illustration, the value of 0.083 (equivalent to 1/12) will serve as a representation for the month of January, while 0.167 (equivalent to 2/12) will serve as a representation for the month of February, and so forth. Subsequently, Australia will be denoted by the numerical value 0.143, which is equivalent to the fraction 1/7. Similarly, Canada will be represented by the numerical value 0.286, corresponding to the fraction 2/7, and so forth. Furthermore, in order to accelerate the process of back propagation learning, the year (ranging from 2014 to 2023) and output data (electric consumption, ranging from 0 to 100.000 GWh) will go through preprocessing, wherein they will be transformed into numerical values ranging from 0 to 1 utilizing the following equation:

$$X_{ij}^{new}=\frac{X_{ij}-\min_l (X_{lj})}{\max_l (X_{lj})-\min_l (X_{lj})}$$

The comprehensive information regarding the input and output of the data can be observed in the provided link: https://bit.ly/datamachinelearningyanuar.

The dataset contains a total of 791 entries, covering electrical consumption data of seven countries from January 2014 to May 2023. The training dataset will consist of eight years of data, specifically from 2014 to 2021, totaling 672 data points for training. The remaining 119 data points will be allocated for testing purposes. As stated in the preceding chapter, prior to being utilized as input for the neural network, the data will undergo preprocessing.

2) Training the Neural Network: The process of training neural networks involves the adjustment of all weights and biases within the neural network, which is determined by the gap between the output of the neural network and the desired output. The given neural network consists of 77 weights and 15 biases, which will undergo training. There are numerous approaches that can be employed, such as:

1. Classic backpropagation
2. Levenberg-Marquardt backpropagation
3. Bayesian Regularization backpropagation
4. Scaled conjugate gradient backpropagation

The initial approach will not be used as it is typically reserved for addressing simple problems. The second method is widely regarded as the most efficient training approach. while the fourth method is known for its minimal computational memory requirements. However, given the complexity and uniqueness of this regression problem, the approach of Bayesian Regularization backpropagation will be employed. To perform weight updates, the Bayesian Regularization backpropagation algorithm employs the following equation.

$$w^{k+1}=w^k-[J^T J+\lambda I]^{-1} J^T e$$

where J is the Jacobian matrix formed by the first derivatives of the network errors e with respect to network weights. λ denotes the Levenberg's damping factor and JTe is the error gradient, which needs to be close to zero at end of the training.

Following the completion of training and the calculation of all weights and biases, the model can be used to forecast testing data values (particularly for the years 2022 and 2023) and compute testing error. Moreover, it can also be utilized for projecting the electrical consumption for 2024.

3) Training the Linear Regression: The training process of linear regression involves the determination of the 9x1 gradient matrix (θ matrix). In this project, the utilization of a fundamental equation for determining the gradient will be employed, as depicted by the equation provided below.

θ=(x^T x)^(-1) (x^T y)

$$\begin{bmatrix}\theta_0\\\theta_1\\\theta_2\\\theta_3\\\theta_4\\\theta_5\\\theta_6\\\theta_7\\\theta_8\end{bmatrix} = \left(\begin{bmatrix}x_{1.1} & x_{2.1} & \cdots & x_{672.1}\\ x_{1.2} & x_{2.2} & \cdots & x_{672.2}\\ \vdots & \vdots & \ddots & \vdots\\ x_{1.9} & x_{2.9} & \cdots & x_{672.9}\end{bmatrix}\begin{bmatrix}x_{1.1} & x_{1.2} & \cdots & x_{1.9}\\ x_{2.1} & x_{2.2} & \cdots & x_{2.9}\\ \vdots & \vdots & \ddots & \vdots\\ x_{672.1} & x_{672.2} & \cdots & x_{672.9}\end{bmatrix}\right)^{-1}$$

$$* \begin{bmatrix}x_{1.1} & x_{2.1} & \cdots & x_{672.1}\\ x_{1.2} & x_{2.2} & \cdots & x_{672.2}\\ \vdots & \vdots & \ddots & \vdots\\ x_{1.9} & x_{2.9} & \cdots & x_{672.9}\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\\\vdots\\y_{672}\end{bmatrix}$$

The θ matrix that has been acquired can subsequently be utilized for predicting the values of the testing data (specifically for the years 2022 and 2023), to perform the calculation of the testing error. Additionally, it can also be employed for forecasting future values for the year 2024.

4) Training Platform: The training and testing processes for both the neural network and linear regression will be conducted using MATLAB Software version R2022a. The initial installation of the neural network can be executed using the "fitnet" command, while the subsequent training of the neural network will be carried out using the "train" command. The gradient matrix in linear regression can be calculated using the standard matrix equation.

C. Testing Strategy

Once the finalization of the model has been completed, for both the neural network and linear regression, the subsequent step involves conducting tests to determine the error associated with each model. Utilized using eight-year data from 2014 to 2021 (672 data), The model will subsequently undergo testing utilizing the dataset encompassing the time period from January 2022 to May 2023 (119 data)

1) Testing the Neural Network: The neural network will receive input consisting of 119 data points representing seven countries, spanning from 2022 to May 2023. The neural network will generate a sequence of 119 numerical values, ranging from 0 to 1, which correspond to the electrical consumption of a specific time period and country. Subsequently, that particular numerical value will undergo a denormalization process in order to ascertain its actual value in gigawatt-hours (GWh). The calculated value will subsequently be compared to the output data in order to determine the level of error. The evaluation of the testing will be conducted utilizing the Root Mean Square Error (RMSE) and the correlation coefficient (R).

2) Testing the Linear Regression: Testing the linear regression is simply perform the matrix calculation of following formula.

$$\begin{bmatrix}y_1\\y_2\\y_3\\\vdots\\y_{119}\end{bmatrix} = \begin{bmatrix}x_{1.1} & x_{1.2} & x_{1.3} & x_{1.4} & x_{1.5} & x_{1.6} & x_{1.7} & x_{1.8} & x_{1.9}\\ x_{2.1} & x_{2.2} & x_{2.3} & x_{2.4} & x_{2.5} & x_{2.6} & x_{2.7} & x_{2.8} & x_{2.9}\\ x_{3.1} & x_{3.2} & x_{3.3} & x_{3.4} & x_{3.5} & x_{3.6} & x_{3.7} & x_{3.8} & x_{3.9}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ x_{119.1} & x_{119.2} & x_{119.3} & x_{119.4} & x_{119.5} & x_{119.6} & x_{119.7} & x_{119.8} & x_{119.9}\end{bmatrix}\begin{bmatrix}\theta_0\\\theta_1\\\theta_2\\\theta_3\\\theta_4\\\theta_5\\\theta_6\\\theta_7\\\theta_8\end{bmatrix}$$

Similar to the neural network, the y1 to y119 will represent electrical consumption in specific time and country, ranging from 0 to 1. The resulting output will also undergo denormalization in order to determine its actual value in gigawatt-hours (GWh), and will subsequently be compared to the actual electrical consumption data. The evaluation of the testing will be conducted utilizing the Root Mean Square Error (RMSE) in percent and the correlation coefficient (R).

III. RESULTS AND DISCUSSION

The training process has been conducted through MATLAB software, and the model has been obtained, as depicted in Figure 3.

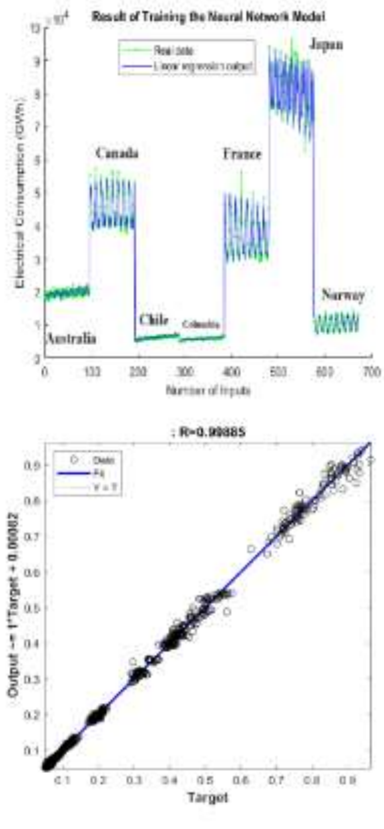

Figure 3 (a) Regression ability of NN model in training process (b) correlation coefficient of training process of NN

Figure 3 (a) depict that after the training process, the model demonstrates a reasonable accuracy to track the expected output trend of seven different countries in 8 years. The performance of the training has been evaluated using RMSE and correlation coefficient. The RMSE of the training process is 4.29% and the correlation coefficient is 0.99885. The model then is used to predict the testing value (119 data), and figure 4 shows the result

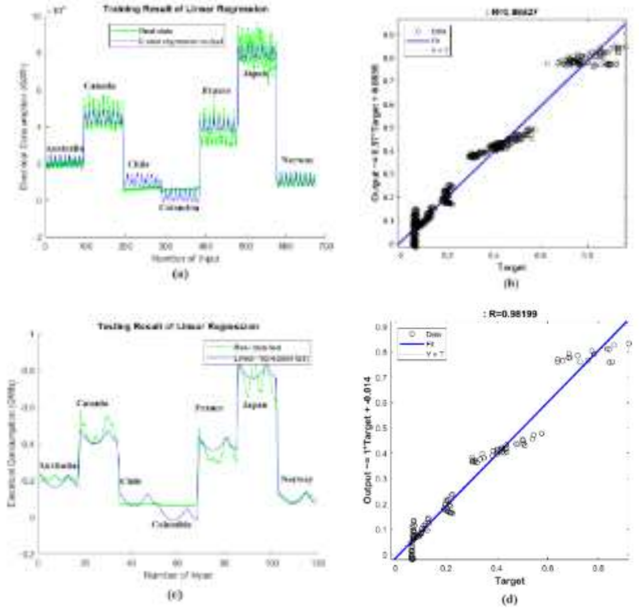

Figure 4 (a) Regression ability of linear regression model in training process

The linear regression model exhibits root means square error (RMSE) values of 41.51% and 40.55% for the training and testing datasets, respectively. The correlation coefficient for the training data is 0.98527, while the correlation coefficient for the testing data is 0.98199. The previously mentioned numerical value is regarded as indicative of poor performance. This can be attributed to the failure of the linear regression model to capture the unique trends exhibited by each individual country. Linear regression is limited in its ability to generalize patterns, as it can only capture and represent them as a single general trend.

TABLE 1
PERFORMANCE COMPARISON BETWEEN NEURAL NETWORK AND LINEAR REGRESSION

Model	Training		Testing	
	RMSE (%)	R	RMSE (%)	R
Neural Network	4.29	0.99885	6.49	0.99727
Linear Regression	41.51	0.98527	40.55	0.98199

Based on the findings presented in Table I, it can be inferred that the neural network model exhibits superior performance in predicting the electrical consumption of seven countries compared to the linear regression model. The neural network successfully captured the unique characteristics of electrical load patterns across different countries, while the linear regression model proved inadequate in this regard. Figure 6 illustrates the projected load consumption in various countries for the year 2024, applying both neural network and linear regression models.

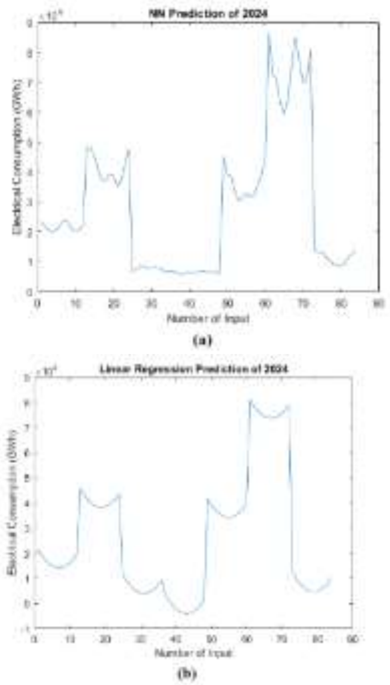

Figure 6 Electrical consumption forecast of 2024 from (a) NN and (b)

must be positioned in the top or in the bottom of the page. The figure is not bordered outside the figure area.

IV. CONCLUSION

The neural network model success to perform electrical forecast with 4.29% training error and 6.49% testing error. The correlation coefficient of this model is 0.99885 and 0.99727 for training and testing, respectively. The linear regression model demonstrates efficacy in predicting electrical outcomes, achieving a training error rate of 41.51% and a testing error rate of 40.55%. The correlation coefficient for the training and testing of this model are 0.98527 and 0.98199, respectively. The efficacy of neural networks outweighs that of linear regression. The neural network effectively captured the unique characteristics of electrical load patterns across various countries, whereas the linear regression model demonstrated insufficiency in this aspect. The neural network model has demonstrated efficacy in predicting electrical consumption for the year 2024, exhibiting an estimated error rate of 6-7%, which is approximate to the testing error. Linear regression has the capability to forecast electrical consumption in 2024, but it has a high margin of error, particularly in Colombia and Chile. This is because linear regression tends to generalize the pattern of electrical load.lists.

REFERENCES

[1] Ishak, Izzaamirah, Nor Salwati Othman, and Nor Hamisham Harun. "Forecasting electricity consumption of Malaysia’s residential sector: Evidence from an exponential smoothing model." F1000Research 11 (2022): 54.

[2] Ciro Eduardo Bazán Navarro, Víctor Josué Álvarez-Quiroz, James Sampi, Adolfo Alfredo Arana Sánchez,

Does economic growth promote electric power consumption? Implications for electricity conservation, expansive, and security policies, The Electricity Journal, Volume 36, Issue 1, 2023, 107235, ISSN 1040-6190.

[3] Bakare, M.S., Abdulkarim, A., Zeeshan, M. et al. A comprehensive overview on demand-side energy management towards smart grids: challenges, solutions, and future direction. Energy Inform 6, 4 (2023).

[4] Dai, Liuyi, Rui Jia, and Xinran Wang. "Relationship between economic growth and energy consumption from the perspective of sustainable development." Journal of Environmental and Public Health 2022 (2022).

[5] Kosowski P, Kosowska K, Janiga D. Primary Energy Consumption Patterns in Selected European Countries from 1990 to 2021: A Cluster Analysis Approach. Energies. 2023; 16(19):6941.

[6] Corazza M, Conti V, Genovese A, Ortenzi F, Valentini MP. A Procedure to Estimate Air Conditioning Consumption of Urban Buses Related to Climate and Main Operational Characteristics. World Electric Vehicle Journal. 2021; 12(1):29

[7] Hasan Rafiq, Prajowal Manandhar, Edwin Rodriguez-Ubinas, Juan David Barbosa, Omer Ahmed Qureshi, Analysis of residential electricity consumption patterns utilizing smart-meter data: Dubai as a case study, Energy and Buildings, Volume 291, 2023, 113103, ISSN 0378-7788.

[8] Madhukumar, Mithun & Sebastian, Albino & Liang, Xiaodong & Jamil, Mohsin & Shabbir, Md. Nasmus Sakib Khan. (2022). Regression Model-Based Short-Term Load Forecasting for University Campus Load. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3144206.

[9] Rui Wang, Hongguang Yun, Rakiba Rayhana, Junchi Bin, Chengkai Zhang, Omar E. Herrera, Zheng Liu, Walter Mérida, An adaptive federated learning system for community building energy load forecasting and anomaly prediction, Energy and Buildings, Volume 295, 2023, 113215, ISSN 0378-7788.

[10] Zongxi Jiang, Luliang Zhang, Tianyao Ji, NSDAR: A neural network-based model for similar day screening and electric load forecasting, Applied Energy, Volume 349, 2023, 121647, ISSN 0306-2619.

[11] Guo-Feng Fan, Ying-Ying Han, Jing-Jing Wang, Hao-Li Jia, Li-Ling Peng, Hsin-Pou Huang, Wei-Chiang Hong, A new intelligent hybrid forecasting method for power load considering uncertainty, Knowledge-Based Systems, Volume 280, 2023, 111034, ISSN 0950-7051.

[12] Khansa Dab, Nilson Henao, Shaival Nagarsheth, Yves Dubé, Simon Sansregret, Kodjo Agbossou, Consensus-based time-series clustering approach to short-term load forecasting for residential electricity demand, Energy and Buildings, Volume 299, 2023, 113550, ISSN 0378-7788

[13] M. Gilanifar, H. Wang, L. M. K. Sriram, E. E. Ozguven and R. Arghandeh, "Multitask Bayesian Spatiotemporal Gaussian Processes for Short-Term Load Forecasting," in IEEE Transactions on Industrial Electronics, vol. 67, no. 6, pp. 5132-5143, June 2020, doi: 10.1109/TIE.2019.2928275..

[14] The International Energy Agency, "Monthly Electricity Statistics".

[15] J. Son, J. Cha, H. Kim and Y. -M. Wi, "Day-Ahead Short-Term Load Forecasting for Holidays Based on Modification of Similar Days' Load Profiles," in IEEE Access, vol. 10, pp. 17864-17880, 2022, doi: 10.1109/ACCESS.2022.3150344

[16] YM. Safarudin, N Musthafa, A. E. Mohamed, S. Z. Khan, " The Number of Nodes Effect to Predict the Electrical Consumption in Seven Distinct Countries" in Journal of Telecommunication Network (Jurnal Jaringan Telekomunikasi) Vol. 13, No.4 (2023).