International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning Approach Using the GRU-LSTM Hybrid Model for Air Temperature Prediction on Daily Basis

Yuslena Sari[1,5], Yudi Firmanul Arifin[2], Novitasari Novitasari[3], Mohammad Reza Faisal*[4]

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: Air temperature has a rapid change movement every day. Temperature prediction is very important as a proper reference base for decision making and good planning for stakeholders. However, in the time series daily temperature prediction, the right number of input combinations has not been found for high accuracy. To overcome this, we propose a deep learning approach using the Hybrid gated recurrent units (GRU) - long short-term memory (LSTM) model. These two deep learning models are very suitable for time series predictions. It has 2 (two) main advantages, namely: A variety of input scenarios is used to find the most reliable performance. (1) the model eliminates the time series decomposition process by embedding a time layer to achieve efficient predictions, and (2) the model achieves a stronger high-level temporal to produce reliable performance. Performance measurement uses root mean squared error (RMSE), mean absolute error (MAE), and R-Squared (R2). Best RMSE on 15-day input, which is 0.07499. The best result of MAE is with a value of 0.0578 at the input of 15 days. The performance results obtained both RMSE and MAE, the smallest of the 15 experimental scenarios is at the input of 15 days. The results of R2 are in line with the results of RMSE and MAE, namely the input in 15 days produces the best R2 close to 1, which is 0.9937.

Keywords: *GRU, LSTM, performance, prediction, temperature*

1. Introduction

In the last century, the global climate has continued to increase [1]. Reported by the official website of the National Centre for Environmental Information from the National Oceanic and Atmospheric Administration (NOAA), the earth's temperature in 2020 warmed by 0.98 °C compared to the 20th century average and was the second warmest year on NOAA's record from 1880 to 2021.

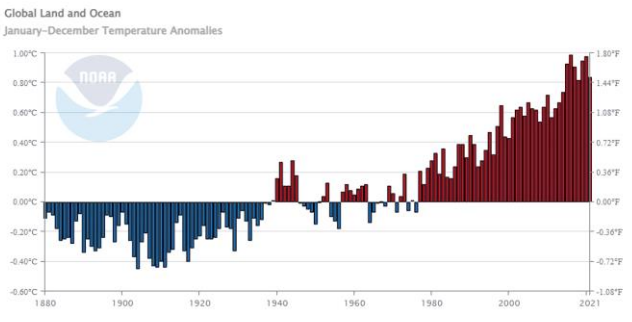

Fig. 2. Comparison between Annual Temperature 1881-2021 and Average Temperature from 1901 until 2000

Reported by the official website of the Indonesian Meteorology, Climatology and Geophysics Agency (BMKG), Indonesia's average monthly temperature in 2022 always increases when compared to Indonesia's average monthly temperature from 1991 to 2020.

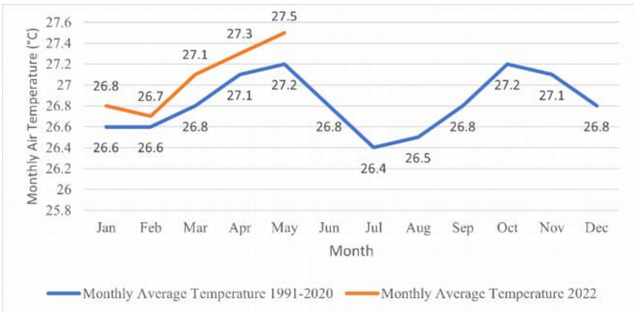

Fig. 1. Air Temperature 1991-2020 and Monthy Air Temperature 2022 Comparison

Changes in temperature affect the rise and fall of sea levels and climate change, temperature changes can also trigger health problems, natural disasters, or extreme weather changes [2]–[6].

Temperature monitoring in Indonesia, especially in South Kalimantan, needs to be done in more detail, because the record for the hottest air temperature in Indonesia was recorded at 40.6°C by the BMKG Banjarbaru Climatology Station on August 16, 1997. To reduce the risk of temperature changes, it is necessary to predict the temperature [5]. Prediction results can be used as a reference for policy formulation by interested parties. Temperature prediction can be done by using time series data [7].

Many studies using time series data for prediction have been carried out, one of which is the use of Recurrent Neural Network (RNN) architecture. Improved RNN architectures such as Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM)

[1] *Doctoral Department of Agricultural Science, Universitas Lambung Mangkurat, INDONESIA*
[2] *Faculty of Forestry, Universitas Lambung Mangkurat, INDONESIA*
[3] *Faculty of Engineering, Universitas Lambung Mangkurat, INDONESIA*
[4] *Department of Computer Science, Universitas Lambung Mangkurat, INDONESIA*
[4] *Department of Information Technology, Universitas Lambung Mangkurat, INDONESIA*
* *Corresponding Author Email: reza.faisal@ulm.ac.id*

have been widely used for prediction and it is concluded that these two methods outperform other prediction techniques and even RNN itself. LSTM is able to overcome the problem of exploding and vanishing gradients caused by long input sequences commonly found in RNNs [8]. LSTM is considered to be able to expose more historical patterns from the time series data used and successfully outperforms other time series prediction methods with significantly improved results [2].

GRU is a simplified version of the LSTM and does not require training time with increased network performance. Compared to LSTM, the computation time of the GRU is shorter and the accuracy is higher [9], [10]. The combination of GRU and LSTM has higher accuracy than the GRU and LSTM models themselves and GRU-LSTM has a lower RMSE value than other prediction methods [11], [12].

Temperature prediction requires a model which has the best performance in order to produce accurate predictions. Evaluation of the predictive model can be done by showing how close the predicted results of the model are to the true value. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R2) have been widely used to evaluate models and can be used to measure predictive model performance.

Time series data need to be compiled into data in the form of series. The data series has a variable number of input parameters. This study will apply one of the Rolling Statistical techniques, namely Exponential (Weighted) Moving Average. Data transformation using Fast Fourier Transform is also carried out. Furthermore, this study will compare the performance results of the GRU-LSTM hybrid model based on the number of input parameters used.

2. GRU-LSTM Hybrid Model for Daily Air Temperature Prediction

This study uses data in the form of air temperature which is processed into various series of data by testing several input parameters, and then through the process of missing value processing and data transformation at the data preprocessing stage. The processed data is then divided into training data and test data and then trained using Deep Learning techniques by applying the GRU-LSTM Hybrid Model. The prediction results of the model will be compared with the test data to obtain Model Performance in the form of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R2) values. In addition, this research will also measure the time of the model training process.

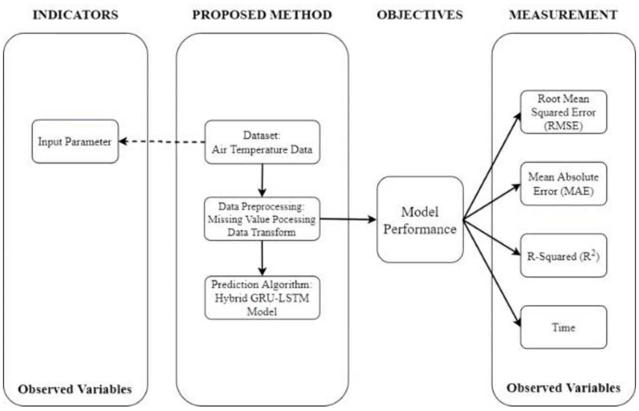

Fig. 3. Framework

2.1. Data

The data used is the average daily air temperature from the Meteorology, Climatology and Geophysics Agency (BMKG) Online Data for the Syamsudin Noor Meteorological Station for 26 years (1 January 1996 – 31 December 2021). The air temperature data are as follows:

Table 1. Datasets

Date	Tavg
01-01-1996	25.60
01-02-1996	25.50
01-03-1996	26.50
01-04-1996	25.70
….	…
12-31-2021	27.20

2.2. Implementation of the GRU-LSTM Hybrid Model

The stages of the GRU-LSTM hybrid model implementation process are as follows:

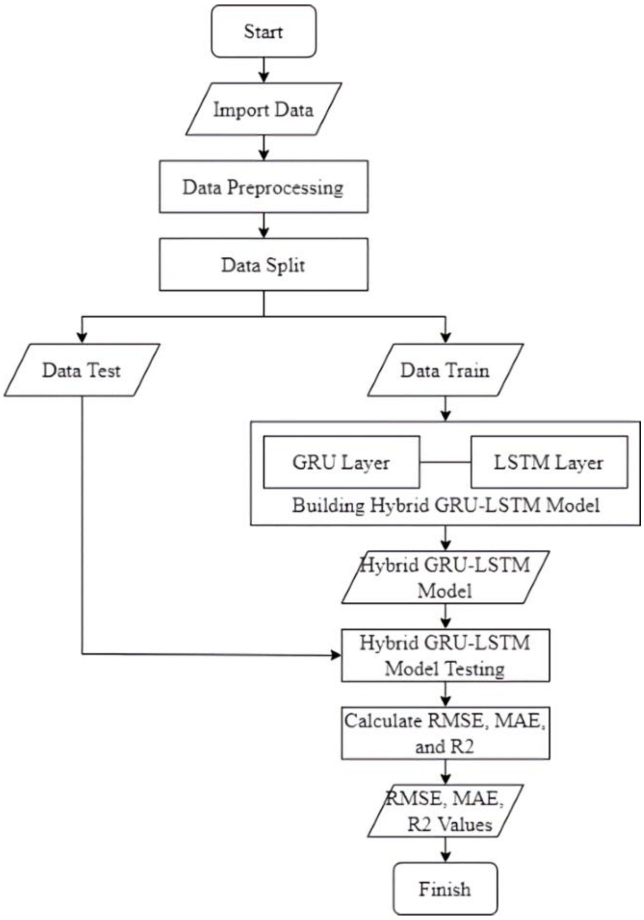

Fig. 4. Model Workflow

2.2.1. Import Data

In this section the daily air temperature data in the form of a file with the extension .csv will be imported using the pandas library.

2.2.2. Data Preprocessing

At the data preprocessing stage, the imported data will be used as a data series first and then use the Exponential (Weighted) Moving Average to fill in the NaN values or empty values and then the data is transformed using the Fast Fourier Transform.

- Exponential (Weighted) Moving Average (EWMA)

EWMA is a statistical method that allows smoothing of datasets by reducing the weight of the data over time and EWMA is sensitive to process shifts [13]. The EWMA standard calculation is as follows:

$$S_t = \lambda e_t + (1-\lambda)S_{t-1} \quad (1)$$

Where S_t is a result of EWMA, t is time stamp, e_t is average output, dan λ is a constant from 0 to 1 which determine the effect of historical data.

- Fast Fourier Trasform (FFT)

FFT is used to change time series data. FFT improves the Discrete Fourier Transform (DFT) algorithm by using periodicity and symmetry, which greatly reduces the number of operations. FFT has the characteristics of simplicity and speed. Therefore, FFT is used to represent errors in the frequency domain to make it easier to identify errors [14]–[16]. FFT can be calculated by [17]:

$$X_t(\mathrm{k}) = \sum_{t=1}^{\bar{N}} x_{ti} e^{\frac{-j2\pi}{N}k(t-1)} \quad (2)$$

where $\boldsymbol{X_i} = [\, \boldsymbol{x_{1i}}\, \boldsymbol{x_{2i}} \cdots \boldsymbol{x_{\bar{N}i}}]^T$, and

i = $\mathbf{1}, \mathbf{2}, \cdots$, **y**, **and** $[\cdot]^T$ represents the transpose from vector [·], $\boldsymbol{k} = \mathbf{0}, \mathbf{1}, \mathbf{2}, \cdots, \bar{\boldsymbol{N}} - \mathbf{1}$.

2.2.3. Splitting Data

The datasets used need to be divided into training/testing subsets for benchmarking. Training sets are used to train the model, and test sets are used to evaluate the model [18]. In this study used a ratio of 90:10. If there are 1000 then 900 data will be the training data used to train the model and 100 data will be the test data used to test the model.

2.2.4. Model Training

The GRU-LSTM hybrid model is built using the GRU and LSTM layers provided by the keras library.

- Gated Reccurent Unit

Gated Recurrent Unit (GRU) is one of the most promising Recurrent Neural Network (RNN) algorithms [11]. In terms of operation, the GRU and LSTM work in the same way but the GRU cell uses a hidden state that combines the forget gate and input gate into an update gate. In addition, GRU also combines hidden and cell states into one state [17]. Therefore, the GRU is referred to as the simplified LSTM variant [9]. The equation used in the GRU is as follows [19]:

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{(t-1)} + b_z) \quad (3)$$

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{(t-1)} + b_r) \quad (4)$$

$$\tilde{h}_t = \tanh(W_h \cdot x_t + (r_t \circ U_h \cdot h_{(t-1)}) + b_h) \quad (5)$$

$$h_t = z_t \circ h_{(t-1)} + (1-z_t) \circ \tilde{h}_t \quad (6)$$

where z_t is update gate, σ sigmoid activation function, W_z is update gate weight, U_z is hidden state weight, x_t is input value (input vector x in timestep t), $h_{(t-1)}$ is previous vector cell value state, b_z is bias update gate, r_t is reset gate, W_r is reset gate weight, U_r is hidden state weight, b_r is reset gate bias, $\tilde{h}_t$ is Output candidate from cell state vector, W_h is cell state vector weight, U_h is hidden state weight, b_h is cell state vector bias, and h_t is cell state vector. The GRU architecture is as follows [9]:

- Long-Short Term Memory

Long-Short Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that allows the network to maintain long-term dependencies between data at a certain time from many timesteps compared to conventional RNNs because LSTM uses special hidden blocks that remember input data for a long time. old [20], [21]. A common LSTM unit consists of a memory cell, forget gate, input gate and output gate, where the purpose of the forget gate is to selectively forget information in the cell state, the input gate decides what new information is stored in the cell state, and the

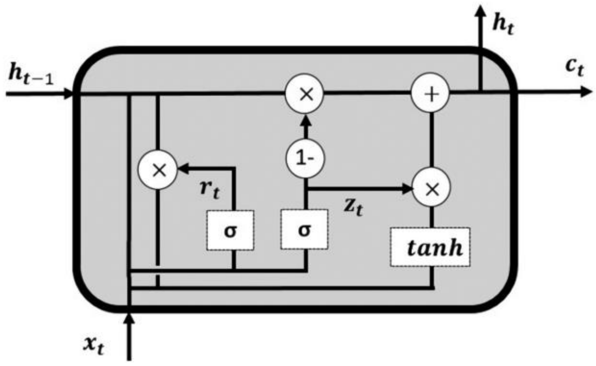

Fig. 5. GRU Architecture

output gate decides what value. that we want to get out. The cell remembers values over variable time intervals and three gates regulate the flow of information into and out of the cell [22]. The process of implementing the LSTM is as follows [23]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

$$c_t = (f_t * c_{t-1} + i_t * \tilde{C}_t) \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t * \tanh(c_t) \quad (8)$$

Where ft is forget gate, σ is sigmoid activation function, Wf is forget gate weight, ht−1 is previous hidden state cell value, xt is input value (input vector x in timestep t), bf is forget gate bias, it is input gate, Wi is input gate weight, bi is input gate bias, $\tilde{C}t$ is candidate gate, Tanh is tanh activation function, Wc is candidate gate weight, bc is candidate gate bias, ct is cell gate, it is input gate, $\tilde{C}t$ is candidate gate, ft is forget gate, ct−1 is previous cell state value, ot is output gate , Wo is output gate weight, bo is output gate bias, ht is hidden state, ct is cell gate. The LSTM architecture is as follows [9]:

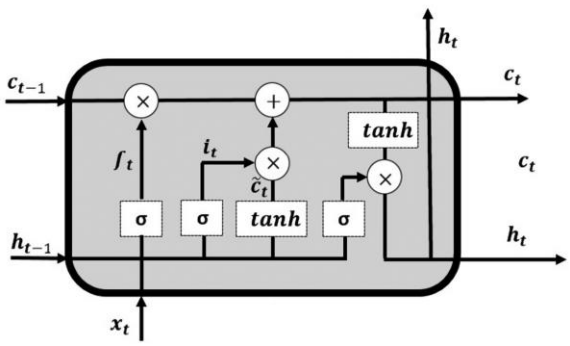

Fig. 6. LSTM Architecture

2.2.5. Model Testing

Model testing is done by utilizing test data to test the GRU-LSTM hybrid model that has been built previously.

2.2.6. RMSE, MAE dan R-Squared Evaluation

The model test results will be used for performance calculations using:

- Root Mean Squared Error (RMSE)

RMSE is one way to test the accuracy of the prediction results by calculating the root of the error value between the predicted value and the actual value [24]. The equation for calculating the RMSE value is [25]:

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

- Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a measure of absolute error between the predicted data and the original data regardless of whether the error is positive or negative [26]. The equation for calculating the MAE value is [25]:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\tilde{y}_i - y_i| \tag{2}$$

- R-Squared (R2)

R-Squared (R2) or commonly known as the coefficient of determination is a statistical measure to study the correlation between the actual data and the predicted data. R2 can have a value between -∞ to 1 where the closer the value of R2 is to 1, the more suitable the existing model with the dataset [11], [24]. The equation for calculating the value of R2 is [25]:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{Y} - y_i)^2} \tag{4}$$

Where n is total data, i is (1, 2, 3, 4, 5, …, l), l represents overall data, $\bar{Y}$ is mean of actual data, y_i is actual value, and $\tilde{y}_i$ is prediction value

2.3. Testing

The test scheme that will be carried out is to test the model using 1 to 15 input parameters. One of the tests will use 7 data per day as input parameters to get 1 target data. The illustration of the data series is as follows:
With d1 = data for the first day, and d(n) = data for the last day

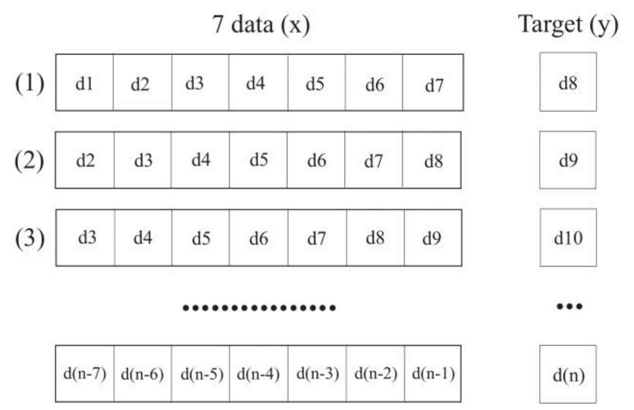

Fig. 7. Series Data Illustration

Therefore, the data from Table 1 will be changed to be the data like shown in Table 2. in this test scheme.

Table 2. Series Data Example

d-7	d-6	d-5	d-4	d-3	d-2	d-1	y
25.60	25.50	26.50	25.70	27.00	25.90	25.10	24.60
25.50	26.50	25.70	27.00	25.90	25.10	24.60	25.70
26.50	25.70	27.00	25.90	25.10	24.60	25.70	27.00
25.70	27.00	25.90	25.10	24.60	25.70	27.00	26.90
….	…	….	…	….	….	…	…
27.00	26.70	27.20	25.40	26.60	25.50	26.80	27.20

And after the data is made into a series for the next process, only the last 9000 data are taken to make it easier to compare the prediction results.

2.4. Result Analysis

The results of model testing will be measured using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R2). The smaller the RMSE and MAE values obtained, the better the performance of the tested model and the closer to 1 R2 value, the better the resulting model. Training Time of all models will also be compared to find out which model has the fastest performance.

3. Result and Discussion

The GRU-LSTM model was built using the python language on the Google Colab platform by utilizing a number of python libraries in the form of time, numpy, pandas, sklearn, tensorflow, and keras. The final 9000 datasets were trained and tested using each test scheme. The GRU-LSTM model has the following architecture [12]:

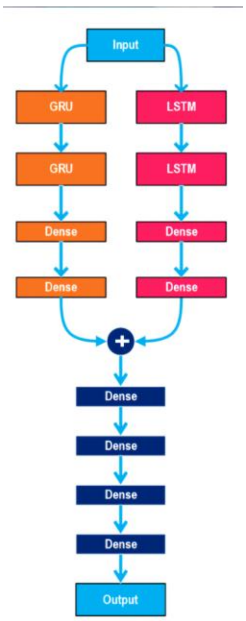

Fig. 8. Hybrid GRU-LSTM Model Architecture

The GRU model is built using 2 GRU Layers and 2 Dense Layers with 240 neuron sizes. The first GRU layer uses ReLU activation, the second GRU layer uses ELU activation, and the second Dense layer uses ReLU activation. The LSTM model is built using 2 LSTM Layers and 2 Dense Layers with 240 neuron sizes. The first LSTM layer uses ReLU activation, the second GRU layer uses ELU activation, and the second Dense layer uses ReLU activation. The GRU model and the LSTM model are combined into a parallel GRU-LSTM model using the add() function from hard.layers. 3 Dense Layers with ELU activation and 240 neuron sizes, as well as 1 Danse Layer with Linear activation and 240 neuron sizes are added to the model. Finally add a Dense Layer with 1 neuron size as the output of the model. The model is built using Adam Optimizer, learning rate is 0.00015, L2 Regularizer is 10-4, epochs is 20, and batch_size is 64. Each test scheme will be trained using this GRU-LSTM hybrid model. The test results are as follows:

Table 3. Model Performance Comparison

Input Parameter	RMSE	MAE	R2	Time (s)
1 day	0.45081751	0.32993866	0.77590259	96.433
2 days	0.22562910	0.16555785	0.94386606	90.817
3 days	0.24706736	0.18409586	0.93269210	90.079
4 days	0.23005558	0.17321041	0.94164194	89.001
5 days	0.20390064	0.15547242	0.95415706	84.512
6 days	0.17912732	0.13597287	0.96461991	84.532
7 days	0.15828254	0.12030653	0.97237506	87.688
8 days	0.14052133	0.10615760	0.97822692	80.318
9 days	0.12531493	0.09481178	0.98268427	89.792
10 days	0.11240998	0.08529736	0.98606699	84.283
11 days	0.10170763	0.07749825	0.98859376	91.876
12 days	0.09417731	0.07221779	0.99022025	89.675
13 days	0.08587602	0.06600229	0.99186834	156.971
14 days	0.07946998	0.06097010	0.99303628	154.703
15 days	0.07499580	0.05774831	0.99379832	107.777

The RMSE and MAE values of the Hybrid GRU-LSTM model using 1 input parameter are very high, when the number of input parameters is added to 2 the RMSE and MAE values decrease significantly. When the input parameters are added to 3 the RMSE and MAE values of the model again increase and decrease again when using 4 input parameters. In the next test, starting from 5 input parameters, 6 parameters, up to 15 parameters the RMSE and MAE values continue to decrease slowly.

The R2 value of the Hybrid GRU-LSTM model using 1 input parameter is quite low, but when the number of input parameters is added to 2, the R2 value increases significantly. When the input parameter is added to 3, the R2 value of the model drops again and increases again when using 4 input parameters. In the next test, starting from 5 input parameters, 6 parameters, up to 15 parameters the RMSE and MAE values continue to decrease slowly.

The duration of model training is not significantly affected by the number of input parameters. In testing using 1 to 6 input parameters, the training duration of the model is getting faster. And again, slowed down when testing 7 input parameters. When using 8 parameter inputs, the model's training duration increases again and slows down again when 9 parameter inputs are used. Likewise, for testing 10, 11 and 12 input parameters, the duration is getting faster, slower and faster again. Then when the input parameters are added to 13 input parameters, the duration of the model training slows down significantly, from 90s to 159s and becomes 154s in the test using 14 parameter inputs and becomes 107s in the last test using 15 input parameters.

4. Conclusion

This study was conducted to see the effect of the number of input parameters on the performance results of the GRU-LSTM hybrid model when predicting air temperature. By changing the number of inputs, the model's performance parameters also change. Increasing the number of input parameters makes the performance model results also increase, but when using 3 input parameters the performance model decreases. The worst model performance is using only 1 input parameter and then 3 input parameters. The best performance model is on 15 parameter inputs, but the duration of model training on tests using more than 12 parameter inputs is greatly increased. The model with the best performance is obtained when using more input parameters. However, model with the best performance and don't take a long training time, it used 12 input parameters in this case.

Acknowlegment

This research was funded by LPDP through the Ministry of

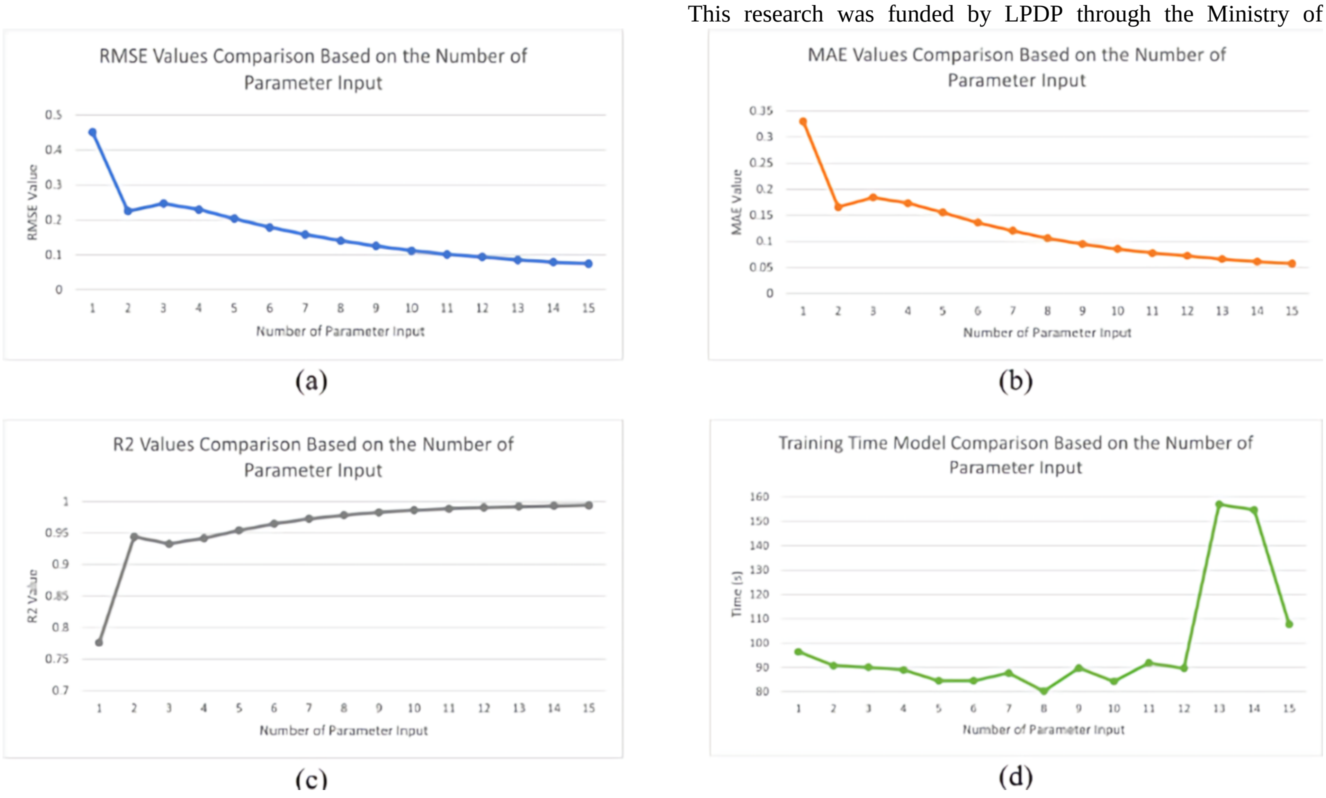

Fig. 9. Model Performance (a) RMSE Values, (b) MAE Values, (c) R2 Values, (d) Training Time

Education, Culture, Research, and Technology Indonesia and the DRTPM grant, National Basic Research Competition scheme, agreement number 113/E5/PG.02.00/PT/2022.

Author Contributions

Conceptualization, Yudi Firmanul Arifin, and Novitasari; methodology, Yuslena Sari; software, Yuslena Sari, and Mohammad Reza Faisal; validation, Yuslena Sari, and Mohammad Reza Faisal; formal analysis, Yudi Firmanul Arifin, and Novitasari; writing— original draft preparation, Yuslena Sari; writing— review and editing, Yudi Firmanul Arifin, Novitasari, and Mohammad Reza Faisal;
All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] M. L. Lin, C. W. Tsai, and C. K. Chen, "Daily maximum temperature forecasting in changing climate using a hybrid of Multi-dimensional Complementary Ensemble Empirical Mode Decomposition and Radial Basis Function Neural Network," J Hydrol Reg Stud, vol. 38, no. September, p. 100923, 2021, doi: 10.1016/j.ejrh.2021.100923.

[2] M. Yu, F. Xu, W. Hu, J. Sun, and G. Cervone, "Using Long Short-Term Memory (LSTM) and Internet of Things (IoT) for Localized Surface Temperature Forecasting in an Urban Environment," IEEE Access, vol. 9, pp. 137406–137418, 2021, doi: 10.1109/ACCESS.2021.3116809.

[3] K. L. Akinwande, A. R. Arotiowa, and A. J. Ete, "Impacts of changes in temperature and exposure time on the median lethal concentrations (LC50) of a combination of organophosphate and pyrethroid in the control of Culex quinquefasciatus, say (Diptera: Culicidae)," Sci Afr, vol. 12, p. e00743, 2021, doi: 10.1016/j.sciaf.2021.e00743.

[4]. Gill, D. R. . (2022). A Study of Framework of Behavioural Driven Development: Methodologies, Advantages, and Challenges. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 09–12. https://doi.org/10.17762/ijfrcsce.v8i2.2068

[5] X. Bi et al., "Impacts of air temperature and its extremes on human mortality in Shanghai, China," Urban Clim, vol. 41, no. June 2021, p. 101072, 2022, doi: 10.1016/j.uclim.2021.101072.

[6]. Ananthakrishnan, B., V. . Padmaja, S. . Nayagi, and V. . M. "Deep Neural Network Based Anomaly Detection for Real Time Video Surveillance". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 54-64, doi:10.17762/ijritcc.v10i4.5534.

[7] D. Cho, C. Yoo, B. Son, J. Im, D. Yoon, and D. H. Cha, "A novel ensemble learning for post-processing of NWP Model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches," Weather Clim Extrem, vol. 35, p. 100410, 2022, doi: 10.1016/j.wace.2022.100410.

[8] Y. Sari, E. S. Wijaya, A. R. Baskara, and R. S. D. Kasanda, "PSO optimization on backpropagation for fish catch production prediction," Telkomnika (Telecommunication Computing Electronics and Control), vol. 18, no. 2, pp. 776–782, 2020, doi: 10.12928/TELKOMNIKA.V18I2.14826.

[9] Z. Liu, Z. Zhu, J. Gao, and C. Xu, "Forecast Methods for Time Series Data: A Survey," IEEE Access, vol. 9, pp. 91896–91912, 2021, doi: 10.1109/ACCESS.2021.3091162.

[10] W. Guo, C. Wu, Z. Ding, and Q. Zhou, "Prediction of surface roughness based on a hybrid feature selection method and long short-term memory network in grinding," International Journal of Advanced Manufacturing Technology, vol. 112, no. 9–10, pp. 2853–2871, 2021, doi: 10.1007/s00170-020-06523-z.

[11] K. E. ArunKumar, D. v. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Forecasting of COVID-19 using deep layer Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) cells," Chaos Solitons Fractals, vol. 146, p. 110861, 2021, doi: 10.1016/j.chaos.2021.110861.

[12]. Kabisha, M. S., Rahim, K. A., Khaliluzzaman, M., & Khan, S. I. (2022). Face and Hand Gesture Recognition Based Person Identification System using Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 105–115. https://doi.org/10.18201/ijisae.2022.273

[13] W. Wu, W. Liao, J. Miao, and G. Du, "Using gated recurrent unit network to forecast short-term load considering impact of electricity price," Energy Procedia, vol. 158, pp. 3369–3374, 2019, doi: 10.1016/j.egypro.2019.01.950.

[14] M. S. Islam and E. Hossain, "Foreign exchange currency rate prediction using a GRU-LSTM hybrid network," Soft Computing Letters, vol. 3, no. June 2020, p. 100009, 2021, doi: 10.1016/j.socl.2020.100009.

[15] E. Haque, S. Tabassum, and E. Hossain, "A Comparative Analysis of Deep Neural Networks for Hourly Temperature Forecasting," IEEE Access, vol. 9, pp. 160646–160660, 2021, doi: 10.1109/ACCESS.2021.3131533.

[16] M. J. Traum, J. Fiorentine. (2021). Rapid Evaluation On-Line Assessment of Student Learning Gains for Just-In-Time Course Modification. Journal of Online Engineering Education, 12(1), 06–13. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/45

[17] Z. Kong, B. Tang, L. Deng, W. Liu, and Y. Han, "Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units," Renew Energy, vol. 146, pp. 760–768, 2020, doi: 10.1016/j.renene.2019.07.033.

[18] S. Mishra, C. Bordin, K. Taharaguchi, and I. Palu, "Comparison of deep learning models for multivariate prediction of time series wind power generation and temperature," Energy Reports, vol. 6, no. September 2019, pp. 273–286, 2020, doi: 10.1016/j.egyr.2019.11.009.

[19] C. Li, Y. Tao, W. Ao, S. Yang, and Y. Bai, "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," Energy, vol. 165, pp. 1220–1227, 2018, doi: 10.1016/j.energy.2018.10.113.

[20] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, "Machine Learning-Based Approach for Hardware Faults Prediction," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 67, no. 11, pp. 3880–3892, 2020, doi: 10.1109/TCSI.2020.3010743.

[21] Y. Fu, Z. Gao, Y. Liu, A. Zhang, and X. Yin, "Actuator and sensor fault classification for wind turbine systems based on fast fourier transform and uncorrelated multi-linear principal component analysis techniques," Processes, vol. 8, no. 9, 2020, doi: 10.3390/pr8091066.

[22] Z. Wu et al., "MoleculeNet: A benchmark for molecular machine learning," Chem Sci, vol. 9, no. 2, pp. 513–530, 2018, doi: 10.1039/c7sc02664a.

[23] X. Liu, Z. Lin, and Z. Feng, "Short-term offshore wind speed forecast by seasonal ARIMA - A comparison against GRU and LSTM," Energy, vol. 227, p. 120492, 2021, doi: 10.1016/j.energy.2021.120492.

[24] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, "Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management," Int J Inf Manage,

vol. 57, no. November, p. 102282, 2021, doi: 10.1016/j.ijinfomgt.2020.102282.

[25] D. Tukymbekov, A. Saymbetov, M. Nurgaliyev, N. Kuttybay, G. Dosymbetova, and Y. Svanbayev, “Intelligent autonomous street lighting system based on weather forecast using LSTM,” Energy, vol. 231, p. 120902, 2021, doi: 10.1016/j.energy.2021.120902.

[26] J. Luo, Z. Zhang, Y. Fu, and F. Rao, “Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms,” Results Phys, vol. 27, p. 104462, 2021, doi: 10.1016/j.rinp.2021.104462.

[27] A. Mellit, A. M. Pavan, and V. Lughi, “Deep learning neural networks for short-term photovoltaic power forecasting,” Renew Energy, vol. 172, pp. 276–288, 2021, doi: 10.1016/j.renene.2021.02.166.

[28] N. AlDahoul et al., “Suspended sediment load prediction using long short-term memory neural network,” Sci Rep, vol. 11, no. 1, pp. 1–22, 2021, doi: 10.1038/s41598-021-87415-4.

[29] D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” PeerJ Comput Sci, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[30] X. Zhang, Q. Zhang, G. Zhang, Z. Nie, Z. Gui, and H. Que, “A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition,” Int J Environ Res Public Health, vol. 15, no. 5, 2018, doi: 10.3390/ijerph15051032.