Short-Term Wind Speed and Temperature Forecasting Model Based on Gated Recurrent Unit Neural Networks

Fahad Radhi Alharbi
Engineering Department
Lancaster University
Lancaster, UK
f.alharbi@lancaster.ac.uk

Denes Csala
Engineering Department
Lancaster University
Lancaster, UK
d.csala@lancaster.ac.uk

***Abstract*: Wind energy generation fluctuations and intermittency issues create inefficiency and instability in power management. The recurrent neural networks (RNNs) prediction approaches are an essential technology that can improve wind power generation and assist in energy management and power systems' performance. In this paper, a prediction model based on Gated Recurrent Unit (GRU) neural networks is proposed to predict wind speed and temperature values one week ahead in the future at hourly intervals. The GRU prediction model automatically learnt the features, used fewer training parameters, and required a shorter time to train compared to other types of RNNs. The GRU model was designed to predict 169 hours ahead as a short-term period of wind speed and temperature values based on 36 years of hourly historical data (1 January 1985 to 6 June 2021) collected from Dumat al-Jandal city. The findings notably indicate that the GRU model has promising performance with significant prediction accuracies in terms of overfitting, reliability, resolution, efficiency, and generalizable processes. The GRU model is characterized by its good performance and influential evaluation error metrics for wind speed and temperature values.**

Keywords— GRU model, recurrent neural network, short-term, update gate, reset gate, prediction, temperature, wind speed.

I. Introduction

Wind power generation has grown exponentially in recent years due to awareness of the global climate change crisis which has required a transition from fossil fuel sources to renewable energy technology. Although it has shown rapid growth in consumption, wind energy's performance is restricted due to the fluctuations in power generation compared to other types of energy generation. The increase in wind power generation creates significant implications for operating systems, such as variations in wind speed and intermittent and non-dispatchable power generation. These factors have an impact on the power generation system's quality, operation protection, distribution effectiveness, and cost [1]. In addition, the unstable nature of wind energy reduces the reliability of the national grid connected to large scale wind farms. Wind energy variations can be counterbalanced by using battery storage technologies. On the other hand, the battery system will play a significant role in raising the cost of wind energy generation and using diesel generators to support the national grid is not decarbonization solution. Forecasting approaches such as neural network deep learning can be employed to predict the future performance of wind speed based on the analysis of the historical data or previous observations. Furthermore, forecasting can reduce energy consumption, balance the demand with supply, and optimize the grid performance [2]. Deep learning forecasting technology has sparked a lot of interest in investigating comparable challenges with common uncertain effects [3]. Moreover, recurrent neural networks (RNNs) are a type of deep learning framework that are effective in interacting with time series. The wind farms' operation can be properly controlled and maintained ahead of time sequences by prediction technology and statistic models. Neural networks are an active research topic in the artificial intelligence field for prediction capabilities with fault-tolerant abilities [2]. In addition, the prediction accuracy was enhanced by neural network deep learning approaches and statistic time sequence models. Nonetheless the main issue with RNNs is the vanishing gradient problem, which restricts the RNN learning of long data sequences [2]. The gradient contains information that is used to update the RNN parameters, and when the gradient declines, the updating of the parameters becomes minimal, indicating that no meaningful learning occurs. The long short-term memory (LSTM) neural network with its complex principles and architecture was introduced to solve the vanishing gradient issue [4], which was proven to have effective performance and outperformed traditional prediction approaches for short-term predictions [5-7]. On the other hand, the challenges are that the prediction models must be fast, efficient, and accurate. In 2014, the first basic model of a gated recurrent unit (GRU) neural network was proposed by Cho et al [8]. The GRU has a simple principle and architecture; is fast, efficient, and more accurate; and has fewer parameters than LSTM. GRU has been shown to perform on smaller and less frequent datasets compared to the LSTM block [2].

In this paper, a prediction model based on a GRU approach is developed for predicting and analyzing the future performance of wind speed and temperature values of short-term hourly intervals (169—h) for one week in advance. In addition, this paper aims to contribute to the issue of high uncertainty and volatility of wind energy, along with the temperature, which can

978-1-6654-3512-3/21/$31.00©2021 IEEE

be used to investigate its impacts on the power generation system influenced by the factors of wind speeds and temperatures. Moreover, the remainder of this paper is organized as follows: The GRU model architecture and method are discussed in Section II. The results of the experiment and discussion are presented in Section III. Conclusions are presented in Section IV.

II. PREDICTIVE MODEL ARCHITECTURE AND ALGORITHM

The characteristics, fundamental principles, and architecture of the GRU neural network prediction model, as well as their algorithms, are discussed in this section. In addition, the collected historical data and the error metrics that are used to evaluate the performance of the GRU model are described in this section.

A. Recurrent Neural Networks (RNN)

GRU is a gate mechanism approach that was developed based on the artificial recurrent neural networks. The GRUs inherit the LSTM's characteristics including several advantages, which include the GRU's lack of an output gate, automatic learning of features, the use of fewer training parameters, and a shorter time required to train in comparison to the other RNN models such as LSTM [9, 10]. In contrast to the LSTM, the GRU can process different sizes of dataset frequencies and time series including a smaller data range. The GRU block consists of two main gates, which are the update gate $z_{(t)}$and the forgetting or reset gate $r_{(t)}$ with an input gate, as illustrated in Fig. 1 [11, 12]. GRUs modulate the data via the gate unit; however, memory access is not provided by a separate memory unit. The update gate is used to regulate the prior state information, which is required for the current state. In addition, the update gate facilitates the prediction model in estimating and evaluating the size of previous data that must be transferred to the future [9]. The higher the value of the update gate, the greater the amount of state information gathered from the previous instant [13]. However, the reset gate is used to govern the state information from the previous instant, which is required to be disregarded. The reset gate primarily controls the amount of activation information that reaches the input as well as how the new input is combined with the previous memory [9]. The lower the value of the reset gate, the greater the amount of information that is ignored.

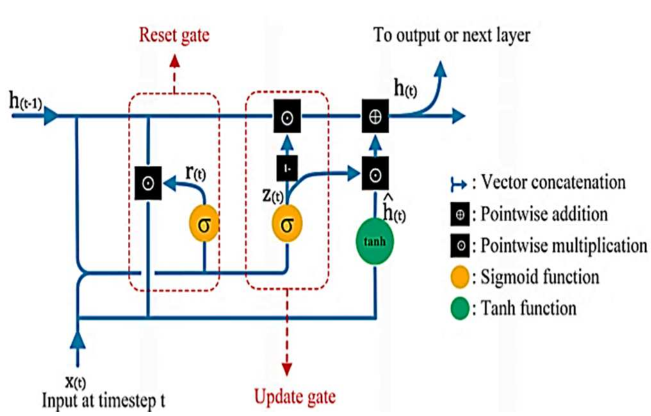

Fig. 1. Architecture of GRU network.

Mathematically, the GRU network can be demonstrated by equations 1− 4, where the $z(t)$ is the update gate, $r(t)$ is reset gate, and $x(t)$ is the input vector at time t [12]. In addition, W and U are the weight parameters matrices that are required to be learned for training. The $h(t-1)$ and $h(t)$ represent the output vector of the past state or previous layer and the current state, respectively. Moreover, b is the base vector, (σ_g) is a sigmoid element, (σ_h) is the hyperbolic tangent,(⊙) is a factor wise multiplication, and the tanh is the activation function that is used to control the flowing values via the network. The extracted features of the (σ_g) and tanh functions are set to be (0, 1) and (1, 1), respectively [11]. The gradient descent technique is used to train the GRU, and the elements are updated continuously until convergence [13].

$$r_{(t)} = \sigma_g(W_r \ x_t + U_r \ h_{t-1} \ + b_r) \tag{1}$$

$$z_{(t)} = \sigma_g(W_r \ x_t + U_z \ h_{t-1} \ + b_z) \tag{2}$$

$$h_{(t)} = 1 - z_t \odot h_{t-1} \ + \ z_t \odot \hat{h}_{(t)} \tag{3}$$

$$\hat{h}_{(t)} = \sigma_h(W_h \ x_t + U_h \ (r_t \ \odot h_{t-1} \ ) + b_h) \tag{4}$$

B. Data Description and Preparation

Dumat al-Jandal city in Al-Jouf region was selected to be a case study for this research. Dumat al-Jandal is located in the northwest of the Saudi Arabia between 29°48′41″ N in latitude and 39°52′06″ E in longitude. A historical dataset was gathered in order to analyze and predict the behavior of wind speed (m/s) at a height of 80m and temperature (°C) at a height of 2m for one week in advance at hourly intervals (169−h). All the real historical data utilized in this research cover the period from 1 January 1985 to 6 June 2021; over 36 years in the historical dataset at hourly intervals are presented in Figs 2−7.

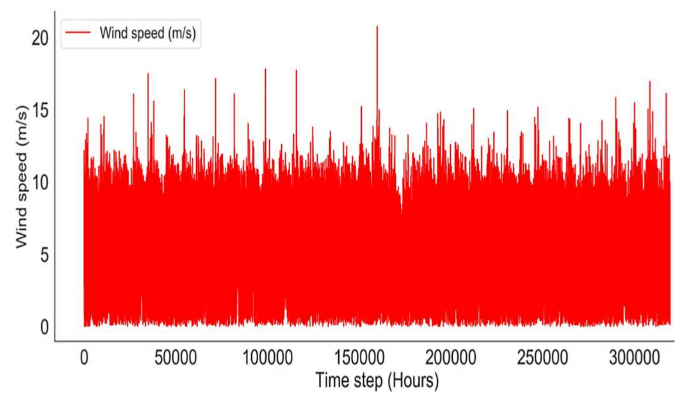

Fig. 2. Historical wind speed performance over 36 years at hourly intervals.

This research used an anemometer to collect the data from the Meteoblue weather service portal [14], which provided the historical wind speeds and temperature data. During the last 36 years, geological conditions have differed and varied over the time, so this large amount of historical data can provide more details, which can be considered by the GRU neural network prediction model. Moreover, the quantity and quality of datasets are a basic factor for GRU models to extract the required features and to obtain an accurate decision for prediction outcomes. Moreover, to guarantee that no values were missed or duplicated, the data were reviewed, validated, and evaluated. The Dickey–Fuller test was applied to ensure that all the data were stationary. Furthermore, the p-values were less than 0.05, and the p-value hypothesis was evaluated (p-values < 0.05).

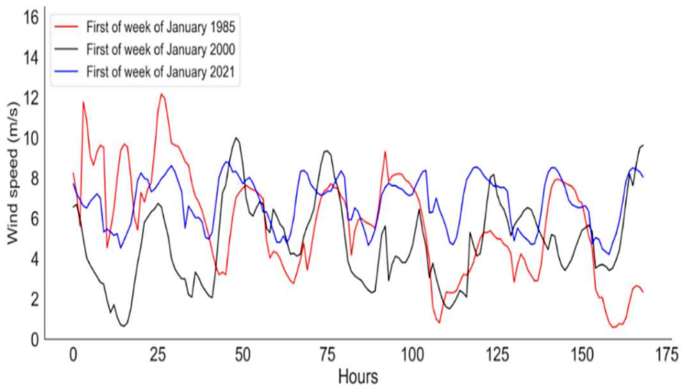

Fig. 3. The first three historical weeks of wind speeds performance during January at hourly intervals.

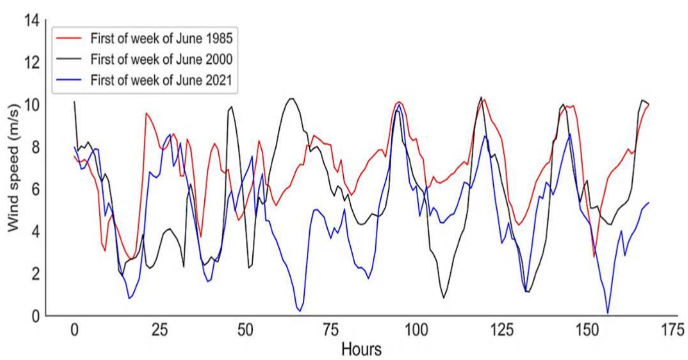

Fig. 4. The first three historical weeks of wind speeds performance during June at hourly intervals.

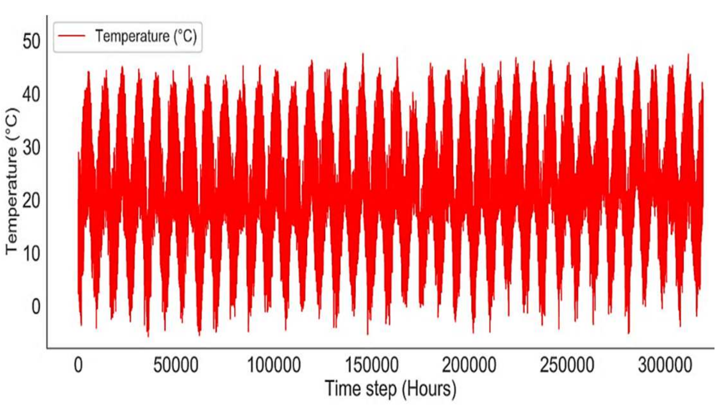

Fig. 5. The historical temperature performance over 36 years at hourly intervals.

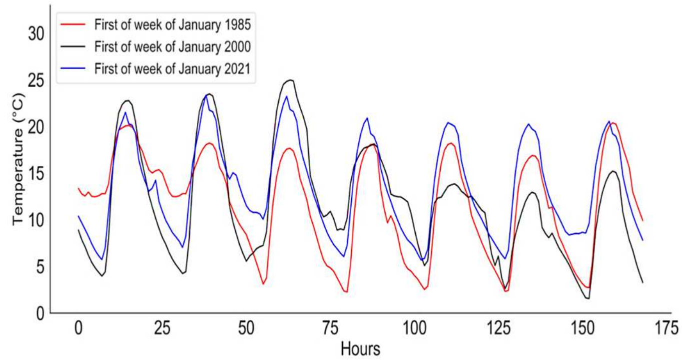

Fig. 6. The first three historical weeks of temperature performance during January at hourly intervals.

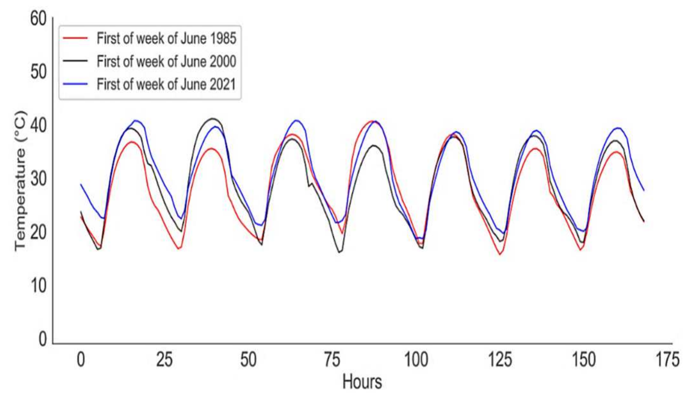

Fig. 7. The first three historical weeks of temperature performance during June at hourly intervals.

C. *Model Architecture*

Python software was used to build and process the codes of the GRU model. Moreover, python programming is widely used to process a large amount of historical data due to its high-level programming language, code readability, and large size of memory function. The historical wind speed and temperature data were screened and cleaned for pre-processing. The GRU-prediction model was developed as a set of coding functions based on the mathematical algorithms (subsection A), including the definition of the model parameters (Fig. 8). The training and testing sets accounted for 70% and 28% of the historical data, respectively, with the remaining data being used to validate the proposed method's predicting capability. The number of neurons in the first and second layers was adjusted to 64. The predicted errors drastically reduced as the number of hidden neurons increased. When the number of neurons was between 50 and 300, the root mean square error, which was caused by the randomization of the input weights for the extreme learning network, remained practically constant with only minor fluctuations. The number of epochs was set to 10, the batch size was set to 64, and the validation split was set to 0.02, which improved the model's fit. Furthermore, in factorization learning networks, the number of inputs and neurons of the hidden layer are the most important parameters for the GRU model.

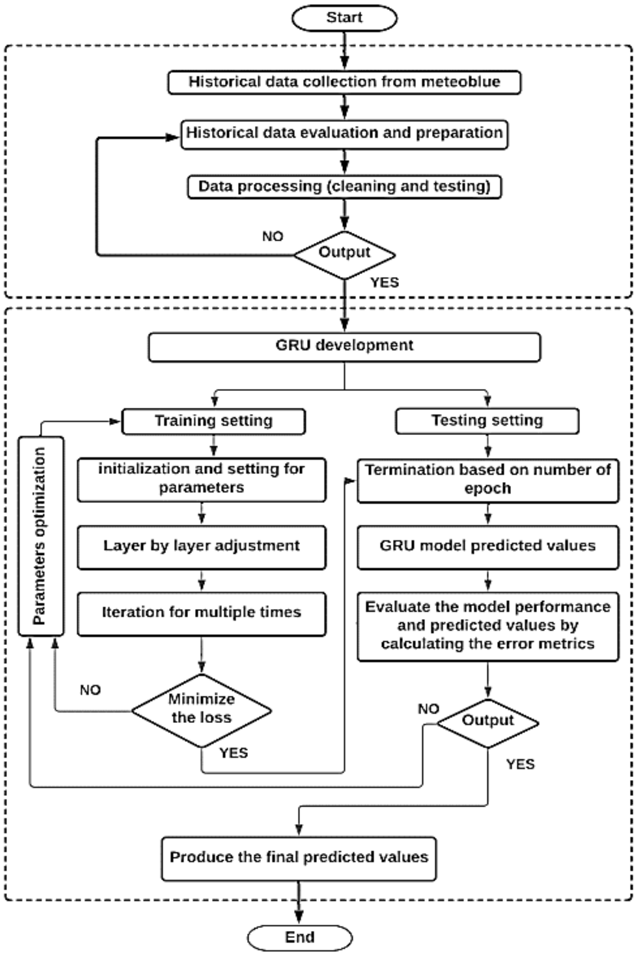

Fig. 8. Flowchart of the GRU model.

D. Prediction Evaluation Criteria

The prediction error is an essential metric for evaluating whether a prediction approach is appropriate for a prediction target. Moreover, the error metrics were used to measure the fitting of the prediction model by comparing the generated values to the historical actual dataset. The correctness and reliability of the GRU prediction model can be examined using three main statistical error measures, which are the mean squared error (MSE), the root mean square error (RMSE), the mean absolute percentage error (MAPE), and the mean absolute error (MAE). Furthermore, the R-squared (R^2) value is a statistical measure that quantifies the amount of a dependent variable's variation adjusted for an independent variable. Whereas the correlation reveals the degree of the association between an independent and dependent variable, the R^2 indicates how the variation relationship of one variable reflects the variation of the other variable. The error metrics that are employed to evaluate the outputs of the GRU prediction model are presented by equations $5-9$. The parameters in these equations are (j), which estimated the data's generic hour; (n), which represents the number of predicted datapoints; (y_j), which is the actual value; and ($y_j^{\wedge}$), which is the forecasted observations value. In addition, the sum of the squared residuals is (SS_{res}), and the absolute number of squares is (SS_{Tot}), which is proportional to the variances in the data.

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}(\mathrm{y_j} - \mathrm{y_j}^{\wedge})^2 \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\frac{(\mathrm{y_j} - \mathrm{y_j}^{\wedge})^2}{\mathrm{n}}} \tag{6}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\left|\frac{\mathrm{y_j} - \mathrm{y_j}^{\wedge}}{\mathrm{y_j}}\right| * 100 \tag{7}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}|\mathrm{y_j} - \mathrm{y_j}^{\wedge}| \tag{8}$$

$$R^2 = 1 - \frac{\mathrm{SS_{res}}}{\mathrm{SS_{Tot}}} \tag{9}$$

III. Results and Discussion

The short-term future performance of the wind speed and temperature of Dumat al-Jandal city were predicted based on a historical dataset covering 36 years using the GRU neural network prediction model. The GRU prediction model's results revealed good performance and noticeable model fitting (see Fig. 9). Despite the enormous quantity of training data, the GRU model did not exhibit overfitting, as seen in Fig. 10. In addition, the wind speed values for 169 hours ahead from 23:00 7 June 2021 to 23:00 14 June 2021 were predicted with significant error metrics, as shown in Fig. 11. The MSE is 0.44 m/s, the RMSE is 0.66 m/s, and the MAE is 0.48 m/s. Moreover, the MAPE is 5%, which highlights the good quality of the GRU model. The R^2 is 93%, which indicates that the variables have a reasonable correlation and variance. The number of variables has a significant impact on the accuracy of the projected models. In addition, the evaluation metrics for all the prediction results are shown in Table I. The epochs size and the error values have a significant correlation, which indicates that, if the number of epochs increases, the errors obviously will decrease, but the training time increases. Furthermore, utilizing a large amount of historical data with a huge quantity of training data resulted in long simulation periods and significant error readings. The GRU prediction model must be processed sequentially, as the succeeding steps are dependent on the previous stages.

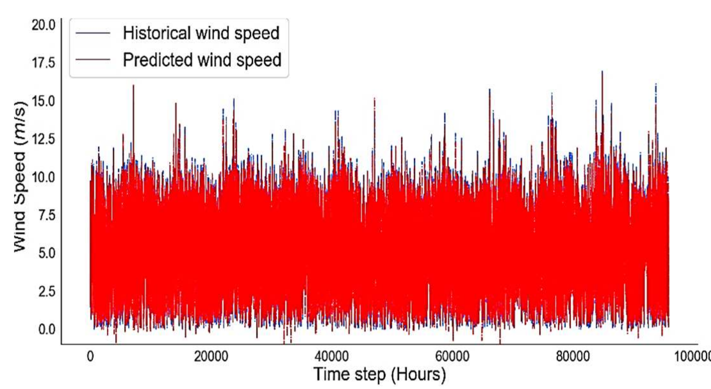

Fig. 9. The actual and predicted wind speed performance.

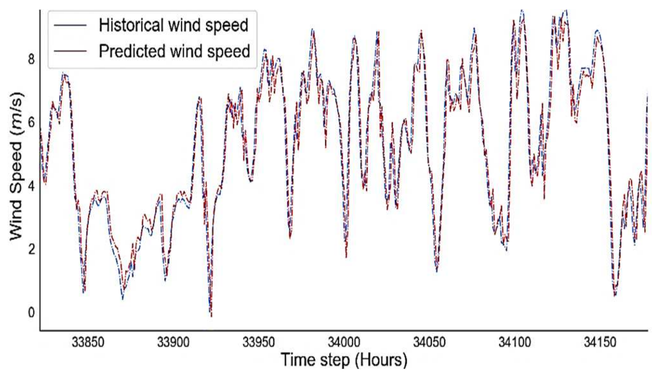

Fig.10. Actual and predicted wind speed performance show the model fitting.

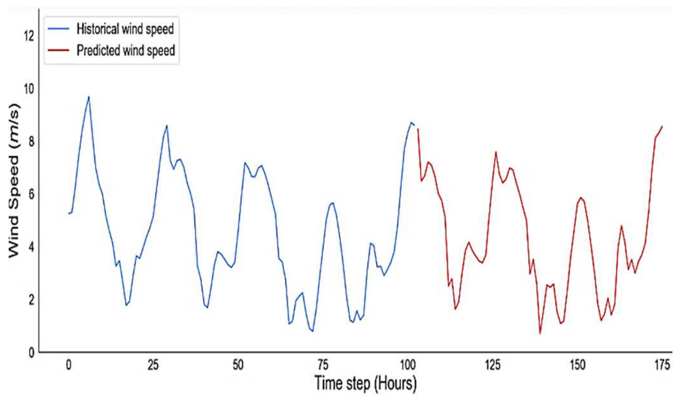

Fig. 11. The future predicted values of the wind speeds for 169 hours ahead.

The GRU model achieved acceptable improvement in the temperature discrimination tasks, as presented in Figs. 12 and 13. The future performance of the temperature values for 169 hours ahead from 23:00 7 June 2021 to 23:00 14 June 2021 were predicted, as presented in Fig. 14. The MSE, RMSE, and MAE error metrics are 0.84 °C, 0.92 °C and 0.68 °C, respectively. The MSE, RMSE, and MAE of the predicted temperature results increased compared to the predicted wind speed error values; however, these error metric values of the GRU model are considered to be relatively small values compared to other prediction models, which demonstrates their flexibility in new dataset observations. In addition, the MAPE and the R^2 were improved which are 2.43% and 99%, respectively. Furthermore, improving this model's generalization capabilities and producing high-accuracy results was one of the most difficult challenges. The variation in model performance when assessing previously observed data, such as training data, over data that the model has never seen before, such as testing data, is referred to as generalizability. Moreover, the model's insufficient generalizability will lead to overfitting of the training data.

In general, the GRU model uses historical time series data to increase prediction accuracy by taking into account the effects of characteristics on anticipated wind speed and temperature values of the next instant. The GRU prediction model does not require the usage of a memory to monitor the transferred information and utilized all the extracted features immediately without any supervision. When a suitable number of historical time series data is selected, the efficiency of the GRU becomes increasingly clear, and the short-term prediction accuracy is developed.

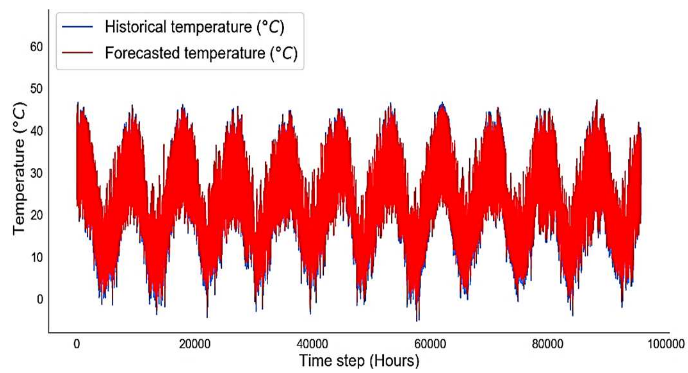

Fig. 12. The actual and predicted temperature performance.

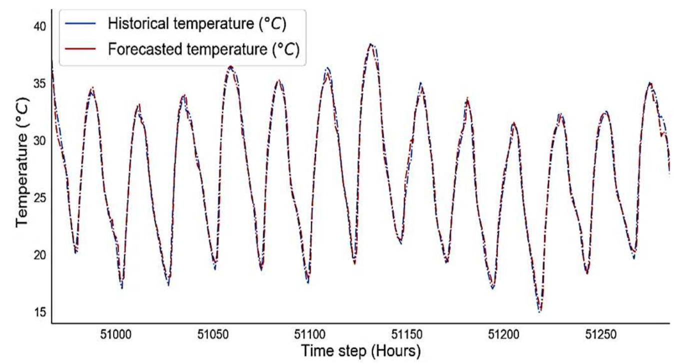

Fig. 13. The actual and predicted temperature performance show the model fitting.

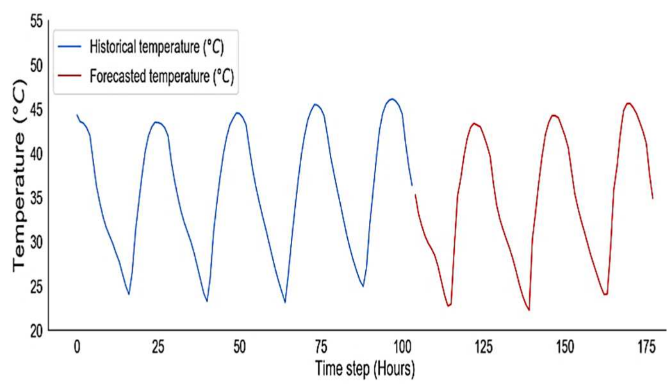

Fig. 14. The future predicted values of the temperature for 169 hours ahead.

TABLE I. THE FORECASTING ACCURACY OF THE GRU MODEL.

Element	**MSE**	**RMSE**	**MAE**	**MAPE (%)**	**P-value (%)**	**R^2 (%)**
Wind speed	0.44 m/s	0.66 m/s	0.48 m/s	5	0.001	93
Temperature	0.84 °C	0.92 °C	0.68 °C	2.43	5.16×10^{-18}	99

IV. CONCLUSIONS

The GRU prediction model was built based on RNNs using the algorithms' strength of neural connections, gated units, and layers, which are the best prediction sequence processes for the future performance of wind speed and temperature. The GRU model was validated to predict $169-\text{h}$ ahead based on a historical dataset covering 36 years from Dumat Al-Jandal city in Saudi Arabia. The hourly intervals from 01 January 1985 to 06 June 2021 were utilized to test the GRU prediction model's ability to cope with this type of historical dataset. The GRU model's underlying architecture provided clearly efficient cell structure, which required fewer training parameters, especially for short-term prediction. The experimental results show that the significant advantages of the GRU model over its competitors are a good degree of prediction accuracies, overfitting, redundancy reduction, and training and testing execution time which showed good performance. The GRU prediction model generated notable error metric values such as MSE, RMSE, MAE, and MAPE, which are 0.44 m/s, 0.66 m/s, 0.48 m/s, and 5%, respectively, for the wind speed values, and 0.84 °C, 0.92 °C, 0.68 °C, and 2.43%, respectively, for the temperature values. Furthermore, the representation of the results demonstrated a remarkable level of the model's learning features, computing efficiency, fast convergences, and capability of GRU approach. However, the model can be improved, and the error metrics can be reduced by enhancing the learning performance, improving the hidden layers, and setting the epoch number or learning iterations. Finally, further study in this area will be expanded and proposed to predict wind speeds by utilizing different sizes of historical datasets and locations including the solar irradiances.

REFERENCES

[1] N. Sodsong, K. M. Yu, and W. Ouyang, "Short-Term Solar PV Forecasting Using Gated Recurrent Unit with a Cascade Model," 2019: IEEE, pp. 292-297.

[2] Y. Wang, M. Liu, Z. Bao, and S. Zhang, "Short-term load forecasting with multi-source data using gated recurrent unit neural networks," Energies, vol. 11, no. 5, p. 1138, 2018.

[3] Y. Zhang and L. Yang, "A novel dynamic predictive method of water inrush from coal floor based on gated recurrent unit model," Natural Hazards, vol. 105, no. 2, pp. 2027-2043, 2021.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[5] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks," 2016: IEEE, pp. 002858-002865.

[6] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," 2017: IEEE, pp. 1-6.

[7] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 841-851, 2017.

[8] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," arXiv preprint arXiv:1406.1078, 2014.

[9] Y.-D. Syu et al., "Ultra-Short-Term Wind Speed Forecasting for Wind Power Based on Gated Recurrent Unit," 2020: IEEE, pp. 1-4.

[10] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," IEEE Transactions on Industrial Electronics, vol. 65, no. 2, pp. 1539-1548, 2017.

[11] Z. Zhang, X. Pan, T. Jiang, B. Sui, C. Liu, and W. Sun, "Monthly and quarterly sea surface temperature prediction based on gated recurrent unit neural network," Journal of Marine Science and Engineering, vol. 8, no. 4, p. 249, 2020.

[12] Y. Wang, W. Liao, and Y. Chang, "Gated recurrent unit network-based short-term photovoltaic forecasting," Energies, vol. 11, no. 8, p. 2163, 2018.

[13] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, "Hybrid deep learning predictor for smart agriculture sensing based on empirical mode decomposition and gated recurrent unit group model," Sensors, vol. 20, no. 5, p. 1334, 2020.

[14] Meteoblue G, Basel, Switzerland. Accessed: June. 06, 2021. [Online]. Available: https://www.meteoblue.com/en/historyplus.