An Ensemble GRU Approach for Wind Speed Forecasting with Data Augmentation

Anibal Flores[1], Hugo Tito-Chura[2]
Grupo de Investigación en Inteligencia Artificial
Universidad Nacional de Moquegua
Moquegua, Peru

Victor Yana-Mamani[3]
E.P. Ingeniería de Sistemas e Informática
Universidad Nacional de Moquegua
Moquegua, Peru

***Abstract*—This paper proposes an ensemble model for wind speed forecasting using the recurrent neural network known as Gated Recurrent Unit (GRU) and data augmentation. For the experimentation, a single wind speed time series is used, from which four augmented time series are generated, which serve to train four GRU sub-models respectively, the results of these sub-models are averaged to generate the results of the proposal ensemble model (E-GRU). The results achieved by E-GRU are compared with those of each sub-model, showing that E-GRU outperforms the sub-models. Likewise, the proposal model (E-GRU) is compared with benchmark models without data augmentation such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), showing that E-GRU is much more precise, reaching a difference of around 15% with respect to the Relative Root mean Square Error (RRMSE) and 11% with respect to the Mean Absolute Percentage Error (MAPE).**

Keywords*—*Wind speed forecasting; recurrent neural networks; gated recurrent unit; ensemble GRU; data augmentation

I. Introduction

Earth's natural greenhouse effect makes life possible as we know [1]. However, human activities, such as the burning of fossil fuels and deforestation, have intensified the natural phenomenon, causing global warming [2], and due to this problem, the exploitation of renewable energies such as solar, wind, thermal energy and others have emerged as excellent alternatives for its solution.

Regarding wind energy, this is harnessed through the use of wind machines or wind motors capable of transforming wind energy into mechanical rotational energy usable for the production of electrical energy. Thus, the prediction of wind speed time series has become an essential task in wind energy farms, this helps in the planning of energy production [3] among others.

In models based on deep learning, the problem of overfitting [4], [5], [6] is usually presented due to the lack of data. Various solutions have been suggested in the literature, such as the use of dropout layers, regularization and data augmentation.

In this work an ensemble model for wind speed forecasting is proposed, it is based on the recurrent neural network known as Gated Recurrent Unit (GRU), where despite having enough historical data for the training phase [7], a data augmentation process is used with the sole objective of improving the precision of the model results, thus it is used the data augmentation technique proposed by Flores et all (2021) "in press" [8]. GRU is used instead of Long Short-Term Memory (LSTM), due to the antecedents such as [9], [10], and others where GRU presents slightly better results than LSTM.

The proposal ensemble model (E-GRU) consists of four GRU sub-models, for which four different augmented time series have been generated from a single wind speed time series. The final result is the average of the four sub-model predictions. The idea of using an ensemble model arises from the need to take advantage of the default and excess predicted values with respect to the observed or original data.

The main contribution of this study is a novel ensemble model (E-GRU) for wind speed time series forecasting based on recurrent neural networks as GRU and data augmentation.

The content of the work has been organized as follows. In the first section, the problem and the respective solution are described. In the second section, the theoretical bases are described, which are the basis of the paper's proposal. In the third section, the methodology followed for the implementation of the proposal is described. In the fourth section, the results achieved are described and discussed. In the last section, the conclusion reached at the end of the study is presented, as well as future work.

II. Background

This section briefly describes some theoretical bases that are important for understanding the content of the paper.

A. Recurrent Neural Networks (RNN)

Just like Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), RNNs are part of the fundamental architectures of Deep Learning, which specialize in working with sequential data, hence their use in natural language processing (NLP) as well as in time series regression.

The best known RNN is probably Long Short-Term Memory (LSTM) known to overcome the vanishing gradient problem in RNNs. Several variants are generated from LSTM, including Gated Recurrent Unit (GRU), which, as mentioned above, for certain case studies, especially in time series, presents better results than LSTM.

The GRU architecture is shown in Fig. 1

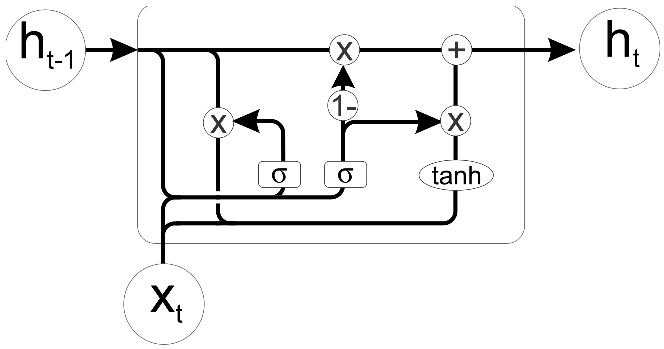

Fig. 1. GRU Architecture.

From Fig. 1, to estimate h_t it is neccesary the following equations:

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

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$\tilde{h}_t = tanh(W x_t + U(r_t \odot h_{t-1}) + b \quad (3)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

Where:

W_z, W_r, W, U_z, U_r, U	Matrices of parameters
b_r, b_z, b	Vectors of parameters
$\boldsymbol{\sigma}$	Element-wise sigmoid function
$\odot$	Element-wise multiplication

B. Data Augmentation

Data augmentation arose to solve overfitting problems in image classification [11] models like CNN and others. Many of these techniques consisted of zooming, rotation, flipping, etc. Later, the concept was transferred to time series classification, here techniques such as time-warping, rotation, scaling, jittering, etc. emerged.

This work uses the technique proposed in "in press" [8] which is based on two basic techniques such as time-warping and jittering. The first one allows to increase the length of the original time series and the second one makes the synthetic data generated with the first one non-linear. Thus, this technique works with two parameters, the block size and the sub-block size, the first indicates the number of synthetic items to insert between each pair of the original time series and the second the number of linear synthetic items in each synthetic block. Fig. 2, shows a graphical view of this data augmentation technique.

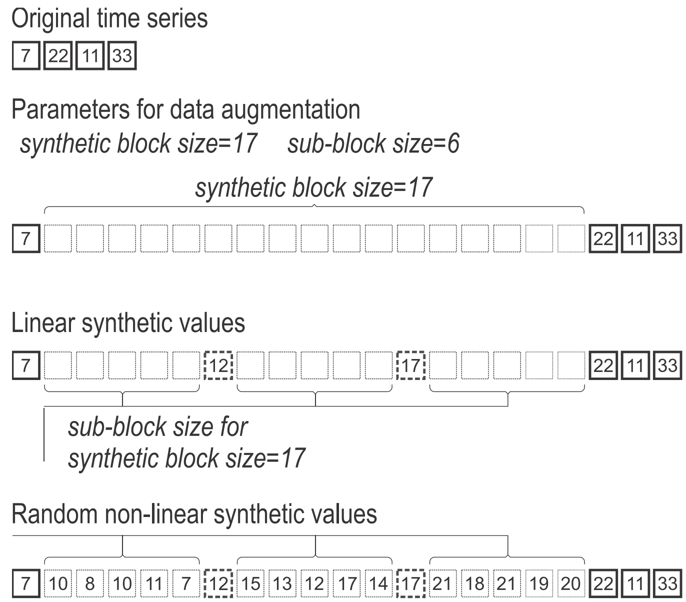

Fig. 2. Data Augmentation based on Time-warping and Jittering [8].

III. Methodology

The methodology followed for the implementation of the proposal is described below.

A. Time Series Selection

The selected daily wind speed time series is the same that was used in the work "in press" [7], and was obtained from the repository of the National Aeronautics and Space Administration (NASA) using Power Data Access Viewer with latitude: -17.6851 and longitude: -71.3515. This corresponds to a point in Ilo city in Peru that has enormous potential for wind energy.

This time series ranges from 1981-01-01 to the present, however, for the purposes of experimentation in this study, the years 1981-2016 will be used for training and the years 2017-2020 for testing.

B. Time Series Imputation

The selected daily wind speed time series does not present NA values, so the application of any data imputation technique was not necessary at this stage.

C. Data Augmentation

In this phase, the data augmentation technique based on time-warping and jittering proposed in [8] was configured according to Table I.

TABLE I. Parameters of Data Augmentation Technique

Time series	Augmented time series	Block-Size	Sub-Block Size	Augmented ítems	Total
1981-2016	TS-1	6	3	78888	92037
	TS-2	6	3	78888	92037
Ítems 13149	TS-3	6	4	78888	92037
	TS-4	6	4	78888	92037

As can be seen in Table I, the first two augmented time series (TS-1 and TS-2) have the same parameters as well as the third and fourth (TS-3 and TS-4), but due to the randomness of the data augmentation technique different items are generated for each synthetic block, this can be seen in Fig. 3.

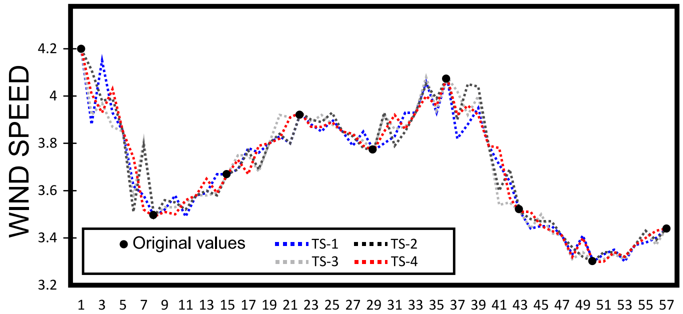

Fig. 3. Augmented Time Series for 9-first Original Items.

D. Ensemble Model Implementation (E-GRU)

At this stage, the ensemble model is implemented. Here the four sub-models have the same characteristics, which are detailed in Table II.

TABLE II. HYPERPARAMETERS OF EACH SUB-MODEL

Sub-Model	Hyperparameters	Values
GRU GRU GRU GRU	Hidden neurons	160
	Epochs	100
	Optimizer	adam
	Drop rate	0.2
	Activation function	ReLu
	Layer 1, 2, 3 y 4	(40,40,40,40)
	Batch size	40

The tools used for implementation of proposal model are Google Colab and tensorflow 2.4.1

E. Evaluation

For the evaluation of the predicted days, it is necessary to extract those corresponding to the original data since these also include predicted synthetic values. For this process, the value of the block-size parameter of the data augmentation technique is considered, which we will call *z*; the predicted time series begins to be traversed and the predicted value located after the *z* value is extracted, then *z* new positions are traversed and the next value is extracted, and so on until reaching the last predicted value.

The model is evaluated through three regression metrics, these correspond to the Root Mean Square Error (RMSE), Relative RMSE (RRMSE) and Mean Absolute Percentage Error (MAPE), which are estimated through equations (5), (6) and (7) respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pi - Oi)^2}{n}} \quad (5)$$

$$RRMSE = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} Oi} * 100 \quad (6)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(O_i - P_i)}{O_i} * 100\right| \quad (7)$$

A graphical version of the proposal model (E-GRU) can be seen in Fig. 4.

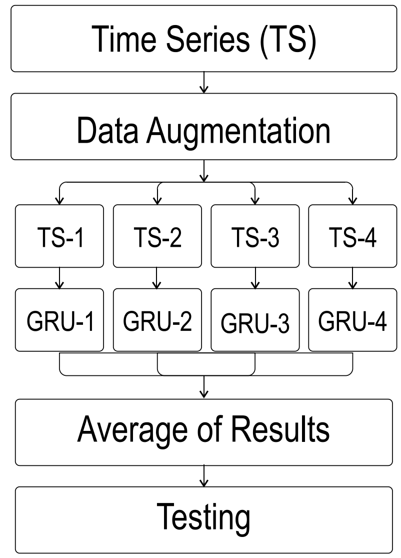

Fig. 4. Proposal Ensemble Model.

IV. RESULTS AND DISCUSSION

After experimentation, this section shows and describes the results achieved.

A. Results

According to Table III and Fig. 5, it can be seen that the ensemble proposal model E-GRU on average surpasses all the sub-models.

Regarding the RMSE, on average E-GRU is superior to all sub-models. However, for the forecast horizon of 500 days, GRU-1 (0.0284) slightly exceeds E-GRU (0.0288), this is the horizon where E-GRU reaches its worst performance.

According to RRMSE on average and in all prediction horizons, E-GRU outperforms all sub-models. It is important to highlight that according to the RRMSE achieved, E-GRU and all the sub-models can be classified as excellent since they present RRMSE <10% [12], [13].

With respect to MAPE, like the previous metrics, on average E-GRU outperforms all sub-models. However, it is important to highlight that GRU-1 for the horizons of 50 and 100 predicted days, manages to surpass E-GRU.

According to Fig. 6, the importance of the ensemble process in the proposal can be appreciated. The data predicted by the sub-models closely approximates the original data by default and excess, and the average operation of the ensemble model makes it much closer to these, making E-GRU more accurate than the sub-models.

Likewise, it is important to highlight the importance of each sub-model, thus in Fig. 6 for the point enclosed in the circle, GRU-4, the worst of the sub-models according to Table III, is the only one that contributes to improving the proposal model precision.

TABLE III. SUB-MODELS VS MODEL RESULTS

Model/ Sub-Model	**Predicted Days**						**Avg**
	50	***100***	***250***	***500***	***1000***	***1461***	
GRU-1							
RMSE	0.0188	0.0235	0.0313	**0.0284**	0.0292	0.0298	0.0268±0.0050
RRMSE	0.5791	0.6854	0.9027	0.8102	0.8389	0.8549	0.7785±0.1297
MAPE	**0.4166**	**0.5031**	0.6041	0.5525	0.5669	0.5759	0.5365±0.0723
GRU-2							
RMSE	0.0299	0.0353	0.0348	0.0331	0.0339	0.0352	0.0337±0.0021
RRMSE	0.9219	1.0269	1.0010	0.9427	0.9723	1.0087	0.9789±0.042
MAPE	0.8045	0.8514	0.8121	0.7050	0.7335	0.7596	0.7776±0.0602
GRU-3							
RMSE	0.0206	0.0320	0.0385	0.0337	0.0349	0.0350	0.0324±0.0067
RRMSE	0.6371	0.9335	1.1091	0.9604	1.0009	1.0019	0.9404±0.1759
MAPE	0.4458	0.6675	0.7851	0.6704	0.6716	0.6774	0.6529±0.1236
GRU-4							
RMSE	0.0511	0.0434	0.0461	0.0419	0.0426	0.0433	0.0447±0.0037
RRMSE	1.5770	1.2656	1.3290	1.1929	1.2234	1.2406	1.3047±0.1537
MAPE	1.2298	1.0035	1.0455	0.9682	0.9930	1.0049	1.0408±0.1053
Proposal Ensemble Model (E-GRU)							
RMSE	**0.0177**	**0.0230**	**0.0255**	0.0288	**0.0238**	**0.0247**	**0.0239±0.0040**
RRMSE	**0.5459**	**0.6713**	**0.7333**	**0.6498**	**0.6839**	**0.7069**	**0.6651±0.0691**
MAPE	0.4375	0.5142	**0.5210**	**0.4622**	**0.4727**	**0.4869**	**0.4824±0.0354**

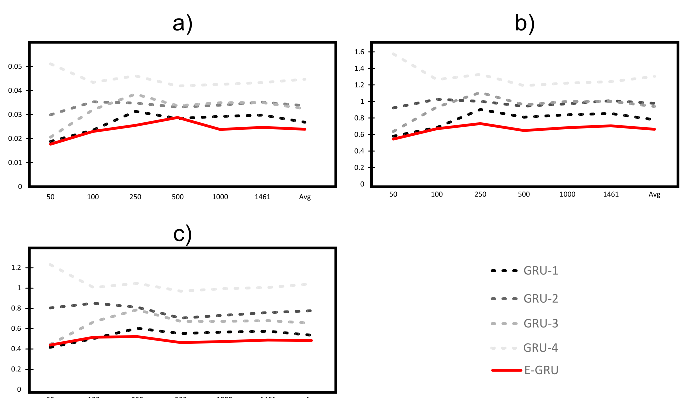

Fig. 5. Comparison of Metrics: Sub Models vs Proposal Model. a) RMSE, b) RRMSE and c) MAPE.

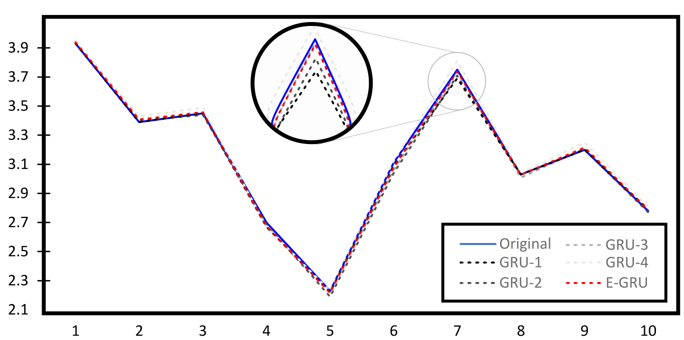

Fig. 6. Comparison of First 10 Predicted Days for Sub Models and Proposal Model.

According to Table IV, in reference to the average and all the prediction horizons it can be seen that the ensemble proposal model E-GRU far exceeds the results of the benchmark models (LSTM and GRU). Here it is important to highlight that the architecture of the LSTM and GRU models is four-layer and use the same hyperparameters as the sub-models of ensemble proposal model, but they do not use data augmentation.

Regarding the RRMSE, there is an average difference of approximately 15% between the results of the ensemble proposal model (E-GRU) and the benchmark models. Likewise, with respect to MAPE, the percentage difference is approximately 11%.

B. Discussion

In this part, the results achieved by the ensemble proposal model E-GRU are compared with those achieved by other state-of-the-art models in the prediction of wind speed time series.

Here, according to Table V, the high precision of the models proposed by Qureshi et al [14] and Flores et al [7] can be highlighted. In the first case, the authors use an architecture based on Deep Neural Networks and Meta Regression with Transfer Learning (DNN MRT), reaching an RMSE = 0.0953. In the second case, the authors use an architecture based on the recurrent neural network GRU including data augmentation, reaching an RMSE = 0.0876.

The E-GRU proposal model uses the same GRU architecture of [7] for each sub-model as well as the same data augmentation technique, the fundamental difference is that instead of using a single augmented time series, it uses four augmented time series, which are different due to the randomness of the technique and also work with different values for the sub-block size parameter.

The results show that the proposal ensemble model manages to surpass the state-of-the-art models including the techniques proposed in [14] and [7].

TABLE IV. BENCHMARK MODELS VS PROPOSAL MODEL RESULTS

Model/ Metric	**Predicted Days**						**Avg**
	50	***100***	***250***	***500***	***1000***	***1461***	
GRU GRU GRU GRU							
RMSE	0.4828	0.5680	0.5761	0.5181	0.5190	0.5146	0.5298±0.0354
RRMSE	14.9025	16.5702	16.592	14.770	14.896	14.744	15.4127±0.907
MAPE	13.0355	14.1669	13.929	12.124	12.314	12.276	12.9745±0.892
LSTM LSTM LSTM LSTM							
RMSE	0.4748	0.5711	0.5824	0.5224	0.5224	0.5380	0.5319±0.0392
RRMSE	14.6557	16.6608	16.772	14.881	14.994	14.843	15.4680±0.973
MAPE	0.4748	13.9701	13.886	12.040	12.164	12.115	10.7751±5.124
Proposal Ensemble Model (E-GRU)							
RMSE	0.0177	0.0230	0.0255	0.0288	0.0238	0.0247	0.0239±0.0040
RRMSE	0.5459	0.6713	0.7333	0.6498	0.6839	0.7069	0.6651±0.0691
MAPE	0.4375	0.5142	0.5210	0.4622	0.4727	0.4869	0.4824±0.0354

TABLE V. COMPARISON WITH RESULTS OF RELATED WORK

Work	Technique	Freq.	Train	Test	RMSE
Zhang et al, 2013 [15]	WTT+SAM+RBFNN	Daily	696	48	0.88
Bokde et al, 2018 [16]	EEMD+PSF	Hourly	2160	720	0.36
Mezaache et al, 2018 [17]	AE+ENN	10-minutes	26000	11000	3.0506
Khodayar et al, 2019 [18]	RBM+IPDL	10-minutes	105120	52560	11.126
Li et al, 2019 [19]	CNN+LSTM	15-minutes	3500	500	3.0012
Liu et al, 2019 [20]	GRU	Daily	811	372	0.9899
Deng et al, 2019 [21]	Bi-GRU			400	6.75
Jiang el at, 2019 [22]	VWC		2304	576	0.2557
Wang et al, 2019 [23]	EWT+KLD	Hourly	14016	3504	1.07
Qureshi et al, 2017 [14]	DNN+MRT	Hourly			0.0953
Yan et al, 2020 [24]	ISSD+LSTM-GOADBN	Hourly	600	100	1.0156
Cheng et al, 2020 [25]	MSSO	10-minutes	2880	720	0.3002
Altan et al, 2020 [26]	DM+LSTM+GWO	10-hours	4397	775	0.1878
Noman et al, 2020 [27]	NARX	10-minutes	Data 2017	Data 2018	0.3590
Luo et al, 2020 [28]	DE+MOO	10-minutes	3200	800	0.2348
Flores et al, 2021 [7]	GRU	Daily	13149	1461	0.0876
Tian et al. 2021 [29]	IWOA-ESN	Hourly	800	200	0.8544
Proposal Model	**GRU**	**Daily**	**13149**	**1461**	**0.0239**

V. CONCLUSION AND FUTURE WORK

According to what is observed in the Results and Discussion section of this paper, it can be concluded that the proposal model allows to improve the results of the state of the art in relation to wind speed forecasting. Likewise, it is important to highlight the importance of the data augmentation process, since all the sub-models implemented for the ensemble proposal model E-GRU present excellent results according to the RRMSE evaluation. Thus, the main advantage of the proposal model with respect to the state-of-the-art models for wind speed prediction is its high precision, and the simplicity of model implementation and each of its respective sub-models.

As a future work, it should be noted that the main weakness of the proposal model lies in the computational cost involved in training 4 GRU models with 92,037 items each. Thus, the minimum amount of synthetic and historical data could be analyzed to obtain satisfactory results. On the other hand, it could be experimented with time series with characteristics different from those of wind speed.

ACKNOWLEDGMENT

To the National University of Moquegua (UNAM), which, through mining canon funds, allowed the execution of the study and publication of its results.

REFERENCES

[1] "What is the greenhouse effect," American Chemical Society, [Online]. Available: www.acs.org. [Accessed 24 06 2021].

[2] B. Bose, "Global warming: energy, environmental pollution, and the impact of power electronics," IEEE Industrial Electronics Magazine, vol. 4, no. 1, pp. 6-17, 2010.

[3] P.J. Zucatelli, E.G.S.. Nascimento, A.Á.B. Santos, A.M.G. Arce, D.M. Moreira, "An investigation on deep learning and wavelet transform to nowcast wind power and wind power ramp: A case study in Brazil and Uruguay," Energy, vol. 230, no. 120842, 2021.

[4] J. Yeomans, S. Thwaites, W.S.P. Robertson, D. Booth, B. Ng, D. Thewlis, "Simulating time-series data for improved deep neural network performance," IEEE Access, vol. 7, pp. 131248-131255, 2019.

[5] G. Forestier, F. Petitjean, H.A. Dau, G.I. Webb, E. Keogh, "Generating synthetic time series to augment sparse datasets," in IEEE International Conference on Data Mining, New Orleans, USA, 2017.

[6] B.K. Iwana, S. Uchida, "Time series data augmentation for neural networks by time warping with a discriminative teacher," arXiv.org, pp. 1-9, 2020.

[7] A. Flores, H. Tito-Chura, V. Yana-Mamani, "Wind speed time series prediction with deep learning and data augmentation "in press"," in Lecture Notes in Networks and Systems, Amsterdam, Netherlands, Springer, 2021.

[8] A. Flores, H. Tito, H. Apaza-Alanoca, "Data augmentation for short-term time series prediction with deep learning," in Lecture Notes in Networks and Systems, London, United Kingdom, Springer, 2021.

[9] A. Flores, H. Tito, D. Centty, "Comparison of Hybrid Recurrent Neural Networks for Univariate Time Series Forecasting," Advances in Intelligent Systems and Computing, p. 375–387, 2021.

[10] J. Zheng, X. Chen, K. Yu, L. Gan, Y. Wang, K. Wang, "Short-term power load forecasting of residential community based on GRU neural network," in International Conference on Power System Technology,, Guangzhou, China, 2018.

[11] C. Shorten, T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 60, pp. 1-48, 2019.

[12] A.N.-L. Huynh, R.C. Deo, D.-A- An-Vo, M. Ali, N. Raj, S. Abdulla, "Near real-time global solar radiation forecasting at multiple time-step horizons using the Long Short-Term Memory network," Energies, vol. 13, no. 14, p. 3517, 2020.

[13] R. Khelifi, M. Guermoui, A. Rabehi, D. Lalmi, "Multi-step ahead forecasting of daily solar radiation components in Saharan climate," International Journal of Ambient Energy, vol. 41, no. 6, pp. 707-715, 2020.

[14] A.S. Qureshi, A. Khan, A. Zameer, A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," Applied Soft Computing, vol. 58, p. 742–755, 2017.

[15] W. Zhang, J. Wang, Z. Zhao, M. Tian, "Short-term wind speed forecasting based on a hybrid model," Jounal of Applied Soft Computing, vol. 92, no. 106294, pp. 1-20, 2013.

[16] N. Bokde, A. Feijoo, K. Kulat, "Analysis of Differencing and Decomposition preprocessing methods for wind speed prediction," Applied Soft Computing, vol. 71, pp. 926-938, 2018.

[17] H. Mezaache, H. Bouzgoud, "Auto-encoder with neural networks for wind speed forecasting," in IEEE International Conference on Communications and Electrical Engineering, El Oued, Algeria, 2018.

[18] M.I. Khodayar, J. Wang, M. Manthouri, "Interval deep generative neural network for wind speed forecasting," IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 3974 - 3989, 2019.

[19] L. Gang; T.F. Wang, F.X. Hu, T.C. Liu, "Algorithm considering multi-dimensional meteorological feature extraction in short-term wind speed prediction," in IEEE Information Technology, Networking, Electronic and Automation Control Conference, Chengdu, China, 2019.

[20] M. Liu, P. Qiu, K. Wei, "Research on wind speed prediction of wind power system based on GRU deep learning," in IEEE Conference on Energy Internet and Energy System Integration, Changsha, China, 2019.

[21] Y. Deng, H. Jia, P. Li, X. Tong, X. Qiu, F. Li, "A deep learning methodology based on bidirectional gated recurrent unit for window power prediction," in IEEE, Xi'an, China, 2019.

[22] P. Jiang, Z. Liu, "Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting," Applied Soft Computing Journal, vol. 82, no. 105587, pp. 1-19, 2019.

[23] J. Wang, Y. Li, "An innovative hybrid approach for multi-step ahead wind speed prediction," Applied Soft Computing Journal, vol. 78, p. 296–309, 2019.

[24] X. Yan, Y. Liu, Y. Xu, M. Jia, "Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition," Energy Conversion and Management, vol. 225, no. 113456, pp. 1-22, 2020.

[25] Z. Cheng, J. Wang, "A new combined model based on multi-objetive salp swarm optimization for wind speed forecasting," Applied Soft Computing Journal, vol. 92, no. 106294, pp. 1-20, 2020.

[26] A. Altan, S. Karasu, E. Zio, "A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer," Applied Soft Computing, vol. 100, no. 106996, 2020.

[27] F. Noman, G. Alkawsi, A.A. Alkahtani, A-S, Al-Shetwi, S.T. Tiong, N. Alalwan, E. Janaka, A.I. Alzahrani, "Multistep short-term wind speed prediction using nonlinear auto-regressive neural network with exogenous variable selection," Alexandria Engineering Journal, vol. 60, no. 1, pp. 1221-1229, 2020.

[28] L. Luo, H. Li, J. Wang, J. Hu, "Design of a combined wind speed forecasting system based on decomposition-ensemble and multi-objective optimization approach," Applied Mathematical Modelling, vol. 89, pp. 49-72, 2021.

[29] Z. Tian, H. Li, F. Li, "A combination forecasting model of wind speed based on decomposition," Energy Reports, vol. 7, pp. 1217-1233, 2021.