

A novel wind power prediction model using graph attention networks and bi-directional deep learning long and short term memory

Ibtissame Mansoury, Dounia El Bourakadi, Ali Yahyaouy, Jaouad Boumhidi

LISAC Laboratory, Department of Computer Science, Faculty of Sciences Dhar-Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Article Info

Article history:

Received May 27, 2023

Revised Jul 9, 2023

Accepted Jul 17, 2023

Keywords:

Bidirectional deep learning

long-short term memory

Graph attention network

Prediction

Renewable energy

Wind power

ABSTRACT

Today, integrating wind energy forecasting is an important area of research due to the erratic nature of wind. To achieve this goal, we propose a new model of wind speed prediction based on graph attention networks (GAT), we added a new attention mechanism and a learnable adjacency matrix to the GAT structure to obtain attention scores for each weather variable. The results of the GAT-based model are merged with the bi-directional deep learning long and short-term memory (BiLSTM) layer to take advantage of the geographic and temporal properties of historical weather data. The experiments and analyzes are carried out using precise meteorological data collected from wind farms in the Moroccan city of Tetouan. We show that the proposed model can learn complex input-output correlations of meteorological data more efficiently than previous wind speed prediction algorithms. Due to the resulting attention weights, the model also provides more information about the main weather factors for the evaluated forecast work.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ibtissame Mansoury

LISAC Laboratory, Department of Computer Science, Faculty of Science Dhar-Mahraz, Sidi Mohamed

Ben Abdellah University

Fez-3000, Morocco

Email: ibtissame.mansoury@usmba.ac.ma

1. INTRODUCTION

Among the many clean and renewable energy sources, wind energy is significant. Wind energy's stochastic and unpredictable character, however, presents considerable difficulties for the power systems' dependable and secure functioning. Wind energy forecasting that is precise and reliable is essential for the integration of wind power into the grid. Numerous atmospheric characteristics are employed in numerical weather prediction (NWP) approaches, which are based on mathematical equations, in the study of weather forecasting [1]. This technique uses equations drawn from the principles of physics and thermodynamics to create predictions. While NWP is an effective solution for many forecasting problems, it still has limitations, such as using an excessive amount of computing power [2]. Additionally, forecasting using NWP may be susceptible to noise that may be included in the observations of meteorological variables because of the assumptions about meteorological variables utilized in this technique [3]. Deep learning machine models can now be trained because of advancements in processing capacity, as an alternative to NWP, deep learning-based models have therefore been utilized more and more lately in weather forecasting tasks [4], [5]. In particular, the approaches based on convolutional neural networks have been utilized in the research on

weather prediction models [3], In addition, there are graph convolutional neural networks [6], and restricted Boltzmann machines [7].

Due to the increase in outstanding results obtained using data-driven approaches, the development of several deep-learning algorithms for weather forecasting problems has been documented in the literature [3], [8]. One way to conceptualize meteorological data is as a time series of successive measurements. The challenges of time series forecasting have often been addressed by recurrent neural networks [9]. For these tasks, bi-directional deep learning long and short term memory (BiLSTM) models are chosen [10], which may impair performance throughout the training period. In contrast, convolutional neural networks (CNNs) are popular and particularly effective for computer vision applications [11]. While vanilla CNNs can handle the geographic aspects of the data, they cannot effectively exploit the temporal properties that are crucial for these weather forecasting applications. Salman *et al.* [3] have developed a 3D-CNN model for forecasting meteorological components. Lara-Benítez *et al.* [12] used vanilla CNNs in conjunction with temporal convolutional networks (TCNs), which is a kind of CNN that utilizes one-dimensional filters to concentrate on the temporal component of the data, to undertake prediction work in the field of meteorology. With the integration of TCN, CNNs can analyze spatiotemporal data. For evaluating graph data, graph neural networks (GNN) [13] have lately grown in prominence. Graph convolutional networks (GCNs) [14] are a kind of GNN that combines the fundamentals of GNNs while maintaining a convolutional architecture by employing graph data rather than assuming that the data is Euclidean. There have been GCN adaptations using a learnable adjacency matrix in [6], although most GCN implementations focus largely on node classification tasks that need knowledge of the graph's nodes and edges. One benefit of this is that normal data can now be represented using an adjacency matrix, which is typically necessary for GNN implementations. However, to extract spatial information, it is desirable to learn the adjacency matrix during the training phase. Due to the transformer design presented in [15], attention methods have lately become quite popular. GNN frameworks have also integrated the attention mechanism. For example, the graph attention networks (GAT) [16] uses a self-attention mechanism similar to a transformer model, to determine the attention coefficient between the vertices of the graph.

Researchers often ignore the impact of spatiotemporal connections between wind turbines in favor of analyzing and modeling the past data for a wind turbine. Therefore, the development of spatiotemporal models with a high prediction of wind energy is crucial. In this article, we propose a novel spatiotemporal wind energy prediction model founded on GAT followed by a BiLSTM algorithm to combine temporal and geographical information of the data. GAT is a subtype of the GNN created specifically to analyze graphical data. It is a model for predicting wind power based on a variety of geographical data. It studies the geographical relationships between different nodes in the graph and how the characteristics of one node can influence the characteristics of another node. It makes optimal use of the attention mechanism to gather data from several nodes, while BiLSTM records temporal correlations between meteorological data collected at distinct periods. Our suggestion to improve the GAT-BiLSTM model makes it possible to calculate the attention of the nodes which produces GAT form in a more accurate and understandable. In our research, past weather data from the wind farm of Tetouan City in Morocco is used to estimate wind speed values for many hours into the future.

This paper is structured as follows: section 2 describes the suggested model. In section 3, you will find a description of the numerical experiments, a discussion of the results, and a comparison to other models. Section 4 contains the conclusion.

2. METHOD

2.1. Graph attention networks (GAT)

The spectral-domain GNN and the spatial-domain GNN are two subcategories of the mainstream GNN. The network topology of the spectral domain GNN is defined by filters, and the spatial domain GNN analyzes the node correlation to construct the network topology. Currently, the Laplace matrix is the only method used by the spectral domain GNN to deconstruct the graph and recover its specific information.

Since the neural network now only decomposes and extracts the characteristic information of the graph using the Laplace matrix, it has disadvantages [17]: i) The noise information in the graph has a simple impact on the eigenvalue, ii) It is challenging to extend the network since the filter it uses can only be applied to a certain structure, and iii) Both the time and financial costs of feature computation and analysis are very high.

GAT presented an attention method to solve the difficulty of GCN [18]. As a result, GAT could give higher weights to important nodes, as neural network parameters and weights are learned simultaneously [19]. At the moment, researchers create the GAT by collecting numerous attention mechanism networks. The feature sets and the transformed feature set of the original nodes constitute the network's input and output for the graph's initial attention layer [20]. The fundamental principle of GAT is to find the spatial link of nodes,

by calculating the weight coefficient of neighboring nodes. The model creates a shared weight matrix W [21] and uses it to examine and determine the coefficient of attention:

$$e_j^i = a(Wh_i, Wh_j) \tag{1}$$

or h denotes the input characteristics, W denotes the weight matrix, and a denotes the attention mechanism.

2.2. Bi-directional deep learning long and short-term memory (BiLSTM)

Hochreiter and Schmidhuber [22] proposed the long-short term memory (LSTM) form of the recurrent neural network (RNN) in 1997, to improve the ability of temporal sequence data to capture long-term connections. It can get around the drawbacks of the vanishing/exploding gradient that standard RNNs experience. Recurrent networks may easily learn across many past time steps without gradient difficulties since the memory of the LSTM helps to maintain the mistake that will be back-propagated via the layers [23]. Blocks are used to store values with varied durations in the LSTM's structure. Each block has two parts that allude to memories: the "Cell state" noted ct , which is notable for being able to recall prior information and displays long-term memory, and the "Hidden state," which only transmits information for the subsequent time step and displays short-term memory.

In contrast, BiLSTM, which consists of two LSTM models, is a powerful and enhanced version of the LSTM processing network. As it comes to the first one, the signal spreads ahead. Whereas in the second, it spreads oppositely. This implies that when processing the material, it concurrently considers both past and future context information [24]. Figure 1 illustrates the construction of the BiLSTM network and demonstrates that each memory block has two LSTM layers. In other words, the output y_t can be determined by adding the values of two opposite hidden (\vec{h}_t and \overleftarrow{h}_t) as you move forward and backward through the layers. The corresponding formulas are provided as [25]:

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \tag{2}$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t+1}) \tag{3}$$

$$y_t = W_{\vec{h}_y} \vec{h}_t + W_{\overleftarrow{h}_y} \overleftarrow{h}_t + b_y \tag{4}$$

where, $LSTM(.)$ displays the LSTM function, b_y is the output layer bias, y_t is the final output of the BiLSTM network, $W_{\vec{h}_y}$ and $W_{\overleftarrow{h}_y}$ are, respectively, the weights of the front and reverse LSTM.

The amount of information sent across the network increases when LSTM is applied twice in each of the two directions. As a result, learning twice as much long-term correlation information in the fundamental architecture of LSTM improves prediction performance [10]. As a result, BiLSTM with dual formation capability vastly outperforms uni-LSTM in a variety of areas.

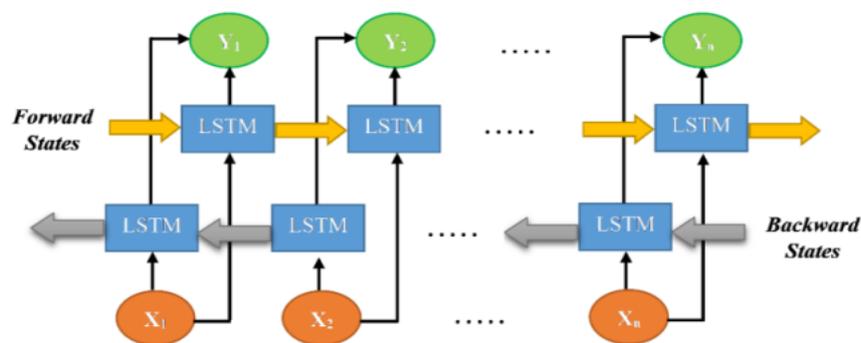


Figure 1. The BiLSTM network structure diagram

2.3. Proposed method

Due to the growing global energy demand, renewable energies are becoming more widespread, especially wind energy. And due to the large instability and fluctuation of wind data, the wind farm is forced to implement practical procedures to manage the wind rotor and increase energy efficiency. Since wind

power forecasting is the basic technique to control and distribute the energy of wind farms, it is more complicated due to the unpredictable and non-linear variation of wind power data. In this article, therefore, we will study wind energy forecasting in detail. The graphical predictor, which is shown in Figure 2, makes up the majority of the GAT set that has been provided.

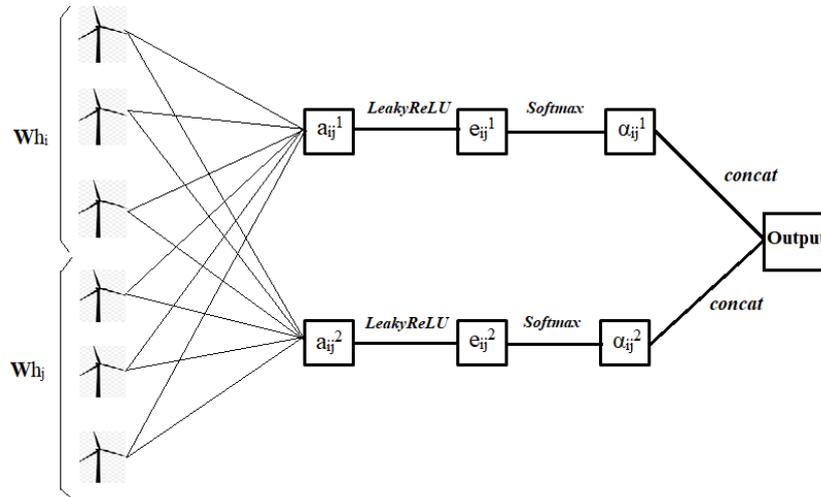


Figure 1. Graph attention network layer

Spatial and temporal wind turbine datasets are pre-processed. The raw wind power data was split between the training set and the testing set. The spatiotemporal wind power prediction model is integrated using GAT followed by BiLSTM as the main predictor, thus, all fundamental weight parameters were defined using the training set. Spatial node feature data from unprocessed datasets are compiled using GAT.

An adjacency matrix is a common way to represent a graph. On the other hand, if the architecture of the graph is unknown, it is possible to determine the weights of the edges of the graph. It has been made learnable in previous works [6], and this has been demonstrated to enhance the capacity of the prediction model. It may also shed light on how the characteristics of one node may be modified by those of other nodes. The typical GAT model requires knowledge of the adjacency matrix of the graph to work [26]. Comparing this model to the one in [26], we make the adjacency matrix learnable. So, we give the network the freedom to fully understand the spatial relationships that exist between the nodes in the graph. To calculate the attention scores for the GAT, this newly acquired information is later utilized. When learning is underway, the adjacency matrix is subjected to the same processes as in (5) [6].

$$\begin{cases} \hat{A} = A + I, \\ \hat{A} = \frac{\hat{A} - \hat{A}_{min}}{\hat{A}_{max} - \hat{A}_{min}}. \end{cases} \quad (5)$$

The identity matrix is therefore added to the adjacency matrix A to take into account the potential self-loop link at each vertex. The updated adjacency matrix A is then normalized using the min-max method. Vertex feature values in the traditional GAT formulation [26] are thought of as being one-dimensional. GAT computes attention ratings based on the vertices and their one-dimensional properties, which are then used to create a fresh depiction of each network node [26]. We presume that the collection of node characteristics is given as $h = \{h_1, h_2, \dots, h_N\}$, or N is the number of nodes. $h_i \in \mathbb{R}^d$ where d is the total number of characteristics found in each node. By using a shared weight matrix, initially, a linear conversion of the attribute values to a specified d' number of attributes. A self-attention approach is then used to calculate the attention scores of each vertex relative to the others, and add to the transformation of the feature values as a dot product. More exactly, the attention score $\alpha_{i,j}$ is calculated in the following manner with the i -th and j -th vertex combination.

$$\alpha_{i,j} = \frac{\exp(\text{LeakyReLU}(a^T[Wh_i||Wh_j]))}{\sum_{k \in N} \exp(\text{LeakyReLU}(a^T[Wh_i||Wh_k]))} \quad (1)$$

In (6) demonstrates the transformation of the characteristics of the i^{-th} and j^{-th} pair of vertex properties are transformed into d' dimensions and concatenated, i.e. $[Wh_i||Wh_j] \in \mathbb{R}^{2d'}$. Here, $h_i \in \mathbb{R}^d$ and $h_j \in \mathbb{R}^d$ consider the feature vectors of the i^{-th} and j^{-th} vertex respectively. $W \in \mathbb{R}^{d \times d'}$ is the associated weight matrix that modifies the properties of each vertex linearly. In GAT, a parameterized single-layer feed-forward neural network models the self-attention process $a \in \mathbb{R}^{2d'}$. A scalar is created by the attention process. The LeakyReLU nonlinearity is then used, with a 0.2 negative slope. The attention score is then normalized with a softmax. The network adjacency matrix determines the neighborhood of the j^{-th} vertex, where all other vertices are taken into account while applying the softmax. In contrast, in our case, the adjacency matrix is learned throughout the learning phase without restricting the connections between the graph's vertices. As a result, in our example, all of the nodes are linked and the adjacency matrix components are used to determine the strength of connections between nodes. The attention scores are multiplied by the linearly changed features of the vertices, and the sum is then calculated over adjacent vertices. Thus, as opposed to the traditional GAT formulation, the combination of the self-attention coefficients a and the concatenated nodes feature $[Wh_i||Wh_j]$ will produce a vector of size d , or the number of weather variables, as opposed to a scalar as it does in the traditional GAT formulation [26]. Subsequently, the following is done to compute the new representation of the i^{-th} vertex:

$$\hat{h}_i = \sigma(\sum_{j \in \mathcal{N}} W h_j \text{diag}(\alpha_{ij}^1, \dots, \alpha_{ij}^d)) \quad (2)$$

where α_{ij}^d is the difference in attention ratings between the i^{-th} and j^{-th} nodes. The final representation of the nodes that we obtain after applying K-independent attention techniques and concatenating their outcomes is as (8).

$$\hat{h}_i = \|\|_{k=1}^K \sigma(\sum_{j \in \mathcal{N}} W^k h_j [\text{diag}(\alpha_{ij}^1, \dots, \alpha_{ij}^d)]^k) \quad (3)$$

The final representation of the i^{-th} node is then determined as (9).

$$\begin{cases} \hat{D}_{ii} = \sum_j \hat{A}_{ij}, \\ \tilde{h}_i = \hat{h}_i \left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} \right) \end{cases} \quad (4)$$

The diagonal matrix D is computed here according to the results of the acquired \hat{A} in (5). The symmetric normalization i.e., $\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}}$, is then done after this. To determine the new feature representation \tilde{h}_i the characteristics of the i^{-th} node given by \hat{h}_i is multiplied by the modified matrix that results. After traversing a spatial layer composed of a flow GAT layer, the output is modified and routed into a BiLSTM network. The GAT layers have been described extract spatial information from meteorological data, Although the BiLSTM is used to enhance the model's capacity to identify temporal correlations in the data and how distinct observations at different time steps impact one another.

3. RESULTS AND DISCUSSION

In this section, we will provide the findings and talk about the numerical analysis. We used information from the wind farm in Tetouan, Morocco, which was gathered on each wind turbine and primarily included the following elements: wind power, wind turbine status, and weather. The original meteorological data used to create this article has been pre-processed, mostly adding the contents of wind speed, wind direction, and ambient temperature. These three categories of meteorological data are essential for creating a very precise spatiotemporal wind power prediction using the proposed GAT-BiLSTM wind forecast model.

Our simulation dataset was divided into training and test sets. As a result, we used 70% of the training phase to introduce the model and 30% of the testing procedure to assess its effectiveness. The suggested approach is put into practice using Python.

To improve and measure the efficiency of the forecasting model during a winter month, we conducted comparative research using the three fundamental algorithms: back-propagation (BP) [27], support vector machine (SVM) [28], and extreme learning machine (ELM) [29], which are the most often used in wind energy forecasting. We included comparison research for a winter month, to investigate the effectiveness of the suggested model and its capacity to perform better throughout the crucial season. We contrasted the results of the BP, SVM, and ELM algorithms, respectively, and the models' investigated observations for a month in the winter, and the prediction results are shown in Figures 3 to 6.

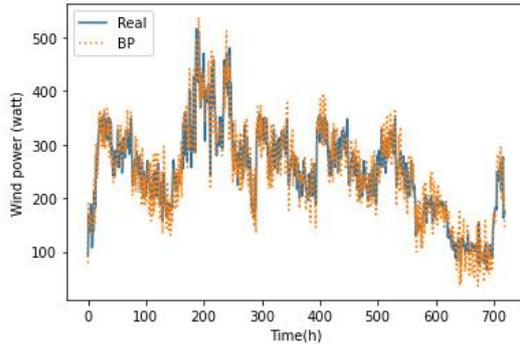


Figure 3. Wind power predicted by BP

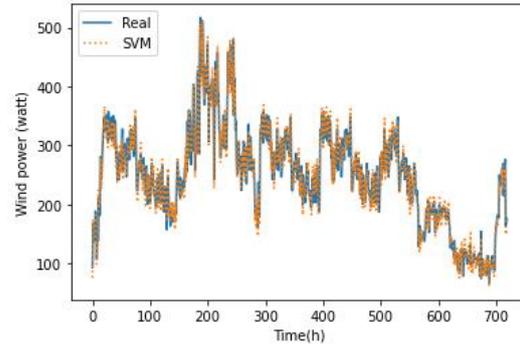


Figure 4. Wind power predicted by SVM

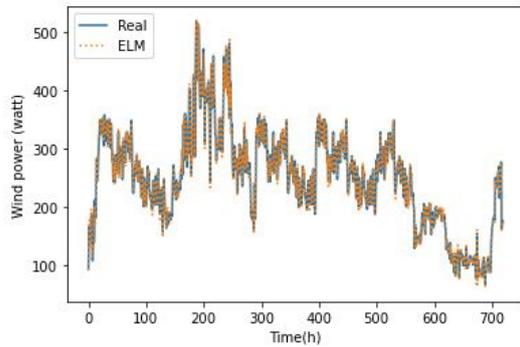


Figure 5. Wind power predicted by ELM

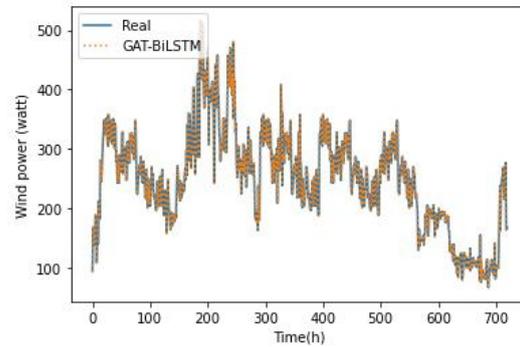


Figure 6. Wind power predicted by GAT-BiLSTM

According to these data, the GAT model provides accurate prediction measures for a month of winter because its curve closely resembles that which was measured. Table 1 lists all forecasting outcomes for a month of winter, based on the four models. The table displays the time convergence of each approach in seconds as well as the MSE score, which can be calculated as (10):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{5}$$

where the number of instances in the test set is n , the predicted output is \hat{y}_i , and the measured output is y_i .

Table 1. Comparison between models for a month in winter

Prediction method	MSE	Time convergence(s)
GAT-BiLSTM	0.1483	0.6582
ELM	0.2267	0.6416
SVM	0.8007	5.7495
BP	0.9135	6.1578

The results are displayed in Table 1 highlighting the execution of the proposed GAT-BiLSTM algorithm concerning the MSE error. Due to their ability to significantly distinguish between test values and projected results of other models, BP and SVM produced the largest MSE. For BP and SVM, the MSEs are 0.9251 and 0.8689 respectively. Also, the ELM error is about 0.3486. All these models have errors greater than 0.1573 produced by GAT-BiLSTM.

In addition, the biggest time convergence value is reached by BP and SVM, which take a long time to compute and need several iterations. On the other hand, ELM and the proposed model give the lowest values, but it still gives a smaller convergence time. Although the suggested GAT-BiLSTM method produced comparatively better predictions than the persistent models and converged quickly, all of these results demonstrated the reliability of the approach.

4. CONCLUSION

The most common renewable energy source must be wind energy for the power system to operate safely, but this is still exceedingly challenging given the volatility of wind speeds and their frequent interruptions. Because of this, predicting wind speed has become crucial for the efficient use of energy. In light of this, we have included a power forecasting model for the production of wind energy in this post. The proposed model attempted to predict the subsequent wind energy generated by the wind turbines using the GAT and BiLSTM deep learning, the so-called GAT-BiLSTM to benefit from historical weather data's spatial and temporal characteristics. The execution of the recommended GAT-BiLSTM algorithm was emphasized by the simulation results since it generated comparatively better predictions than the other examined algorithms. It is an extremely quick, potent, and dynamic learning algorithm. As a consequence, we can conclude that the GAT-BiLSTM may be used quite flexibly in the field of wind power forecasting. However, this model can give more accurate results if we optimize the hyperparameters of the whole model. For this reason, in future works, we think to use an optimization algorithm like a genetic algorithm to improve the prediction precision. Moreover, we also think to enhance our model by using several wind farms in different cities.

REFERENCES

- [1] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, Sep. 2015, doi: 10.1038/nature14956.
- [2] X. Chen, M. Wei, and J. Sun, "Workflow-based platform design and implementation for numerical weather prediction models and meteorological data service," *Atmospheric and Climate Sciences*, vol. 07, no. 03, pp. 337–351, 2017, doi: 10.4236/acs.2017.73025.
- [3] A. G. Salman, B. Kanigoro, and Y. Heryadi, "Weather forecasting using deep learning techniques," in *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Oct. 2015, pp. 281–285, doi: 10.1109/ICACSIS.2015.7415154.
- [4] A. Saeed, C. Li, Z. Gan, Y. Xie, and F. Liu, "A simple approach for short-term wind speed interval prediction based on independently recurrent neural networks and error probability distribution," *Energy*, vol. 238, Jan. 2022, doi: 10.1016/j.energy.2021.122012.
- [5] Z. Tian, H. Li, and F. Li, "A combination forecasting model of wind speed based on decomposition," *Energy Reports*, vol. 7, pp. 1217–1233, Nov. 2021, doi: 10.1016/j.egy.2021.02.002.
- [6] J. Zhao, Z. Yan, X. Chen, B. Han, S. Wu, and R. Ke, "k-GCN-LSTM: a k-hop graph convolutional network and long-short-term memory for ship speed prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 606, Nov. 2022, doi: 10.1016/j.physa.2022.128107.
- [7] R. Hrasko, A. G. C. Pacheco, and R. A. Krohling, "Time series prediction using restricted Boltzmann machines and backpropagation," *Procedia Computer Science*, vol. 55, pp. 990–999, 2015, doi: 10.1016/j.procs.2015.07.104.
- [8] D. Qiao, S. Wu, G. Li, J. You, J. Zhang, and B. Shen, "Wind speed forecasting using multi-site collaborative deep learning for complex terrain application in valleys," *Renewable Energy*, vol. 189, pp. 231–244, Apr. 2022, doi: 10.1016/j.renene.2022.02.095.
- [9] Z. Duan, H. Liu, Y. Li, and N. Nikitas, "Time-variant post-processing method for long-term numerical wind speed forecasts based on multi-region recurrent graph network," *Energy*, vol. 259, Nov. 2022, doi: 10.1016/j.energy.2022.125021.
- [10] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 3285–3292, doi: 10.1109/BigData47090.2019.9005997.
- [11] M. Billah, X. Wang, J. Yu, and Y. Jiang, "Real-time goat face recognition using convolutional neural network," *Computers and Electronics in Agriculture*, vol. 194, Mar. 2022, doi: 10.1016/j.compag.2022.106730.
- [12] P. Lara-Benítez, M. Carranza-García, J. M. Luna-Romera, and J. C. Riquelme, "Temporal convolutional networks applied to energy-related time series forecasting," *Applied Sciences*, vol. 10, no. 7, Mar. 2020, doi: 10.3390/app10072322.
- [13] J. Fox and S. Rajamanickam, "How robust are graph neural networks to structural noise?" *arXiv preprint arXiv:1912.10206*, Jan. 2020.
- [14] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 2018, pp. 499–508, doi: 10.1145/3178876.3186116.
- [15] Z. Chen *et al.*, "Spatial-temporal short-term traffic flow prediction model based on dynamical-learning graph convolution mechanism," *Information Sciences*, vol. 611, pp. 522–539, Sep. 2022, doi: 10.1016/j.ins.2022.08.080.
- [16] X. Wang *et al.*, "Heterogeneous graph attention network," in *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, Mar. 2019, pp. 2022–2032, doi: 10.1145/3308558.3313562.
- [17] H. Peng *et al.*, "Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting," *Information Sciences*, vol. 521, pp. 277–290, Jun. 2020, doi: 10.1016/j.ins.2020.01.043.
- [18] Z. Yang and S. Dong, "HAGERec: hierarchical attention graph convolutional network incorporating knowledge graph for explainable recommendation," *Knowledge-Based Systems*, vol. 204, Sep. 2020, doi: 10.1016/j.knsys.2020.106194.
- [19] C. Jin *et al.*, "HetGAT: a heterogeneous graph attention network for freeway traffic speed prediction," *Journal of Ambient Intelligence and Humanized Computing*, Jan. 2021, doi: 10.1007/s12652-020-02807-0.
- [20] F. Zhou, Q. Yang, T. Zhong, D. Chen, and N. Zhang, "Variational graph neural networks for road traffic prediction in intelligent transportation systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2802–2812, Apr. 2021, doi: 10.1109/TII.2020.3009280.
- [21] F. Wang *et al.*, "Graph attention convolutional neural network model for chemical poisoning of honey bees' prediction," *Science Bulletin*, vol. 65, no. 14, pp. 1184–1191, Jul. 2020, doi: 10.1016/j.scib.2020.04.006.
- [22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [23] D. El Bourakadi, A. Yahyaouy, and J. Boumhidi, "Intelligent energy management for micro-grid based on deep learning LSTM prediction model and fuzzy decision-making," *Sustainable Computing: Informatics and Systems*, vol. 35, Sep. 2022, doi: 10.1016/j.suscom.2022.100709.

- [24] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Computing and Applications*, vol. 33, no. 10, pp. 4741–4753, May 2021, doi: 10.1007/s00521-020-05532-z.
- [25] A. Kulshrestha, V. Krishnaswamy, and M. Sharma, "Bayesian BiLSTM approach for tourism demand forecasting," *Annals of Tourism Research*, vol. 83, Jul. 2020, doi: 10.1016/j.annals.2020.102925.
- [26] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" *arXiv preprint arXiv:2105.14491*, May 2021.
- [27] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Applied Energy*, vol. 87, no. 7, pp. 2313–2320, Jul. 2010, doi: 10.1016/j.apenergy.2009.12.013.
- [28] M. Liu, Z. Cao, J. Zhang, L. Wang, C. Huang, and X. Luo, "Short-term wind speed forecasting based on the Jaya-SVM model," *International Journal of Electrical Power and Energy Systems*, vol. 121, Oct. 2020, doi: 10.1016/j.ijepes.2020.106056.
- [29] D. El Bourakadi, A. Yahyaouy, and J. Boumhidi, "Multi-agent system based on the extreme learning machine and fuzzy control for intelligent energy management in microgrid," *Journal of Intelligent Systems*, vol. 29, no. 1, pp. 877–893, Dec. 2019, doi: 10.1515/jisys-2018-0125.

BIOGRAPHIES OF AUTHORS



Ibtissame Mansoury     obtained a master's degree in Imaging and Business Intelligence from Sidi Mohamed Ben Abdellah University, Morocco, in 2017. Now she is a Ph.D. student at Sidi Mohammed Ben Abdellah University, Morocco, 2017. Her research interests include artificial intelligence and smart grids. She can be contacted at email: ibtissame.mansoury@usmba.ac.ma.



Dounia El Bourakadi     received her Ph.D. degree in 2020 from the University Sidi Mohamed Ben Abdellah (USMBA), Fez, Morocco. Her research interests are in artificial intelligence, optimization, and energy management. She can be contacted at email: dounia.elbourakadi@usmba.ac.ma.



Ali Yahyaouy     is a professor at the Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Morocco. He obtained his Ph.D. degree at the University of Technology of Belfort-Montbéliard, France, and Sidi Mohamed Ben Abdellah University, Morocco in 2010. His research activities mainly concern machine learning, multi-agent systems, smart grids, and ITS. He can be contacted at email: ali.yahyaouy@usmba.ac.ma.



Jaouad Boumhidi     is a Professor of Computer Science at the Faculty of Sciences, Fez Morocco. He received his Ph.D. in Computer Science from The University of Sidi Mohamed ben Abdellah in 2005. His research interests are in machine learning, deep learning, and intelligent transportation systems. He can be contacted at email: jaouad.boumhidi@usmba.ac.ma.