2023

BiLSTM–BiGRU: A Fusion Deep Neural Network For Predicting Air Pollutant Concentration

Prasanjit Dey
Technological University Dublin, Ireland, prasanjit.dey@tudublin.ie

Soumyabrata Dev
Trinity College Dublin, Ireland, soumyabrata.dev@adaptcentre.ie

Bianca Schoen-Phelan
Technological University Dublin, bianca.schoenphelan@tudublin.ie

2023-07-16

BiLSTM–BiGRU: A Fusion Deep Neural Network For Predicting Air Pollutant Concentration

Prasanjit Dey

Soumyabrata Dev

Bianca Schoen-Phelan

IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10282742

BILSTM−BIGRU: A FUSION DEEP NEURAL NETWORK FOR PREDICTING AIR POLLUTANT CONCENTRATION

Prasanjit Dey[1], *Soumyabrata Dev*[2,3] *Bianca Schoen Phelan* [1]

[1]School of Computer Science, Technological University Dublin, Ireland
[2]The ADAPT SFI Research Centre, Dublin, Ireland
[3]School of Computer Science, University College Dublin, Ireland

ABSTRACT

Predicting air pollutant concentrations is an efficient way to prevent incidents by providing early warnings of harmful air pollutants. A precise prediction of air pollutant concentrations is an important factor in controlling and preventing air pollution. In this paper, we develop a bidirectional long-short-term memory and a bidirectional gated recurrent unit (BiLSTM−BiGRU) to predict $PM_{2.5}$ concentrations in a target city for different lead times. The BiLSTM extracts preliminary features, and the BiGRU further extracts deep features from air pollutant and meteorological data. The fully connected (FC) layer receives the output and makes an accurate prediction of the $PM_{2.5}$ concentration. The model is then compared with five other deep learning models in terms of root mean square error (RMSE), mean absolute error (MAE) and correlation (R^2) over different lead times. The results indicate that the proposed model has at least 2.2 times lower RMSE than the other models.

***Index Terms*—** Fusion deep neural network, LSTM, GRU, $PM_{2.5}$ concentration prediction

1. INTRODUCTION

In recent years, increasing health risks from air pollution have become a major problem all over the world [1]. As a result, the prediction of air pollutant concentrations has received a lot of attention because it plays an important role in reducing air pollution and managing the environment [2, 3]. Several researchers have recognized that there is a major issue in predicting air pollutant concentrations, as there is no consistent pattern in the historical pollutant data.

China has produced the highest level of $PM_{2.5}$ in the world. According to the study, $PM_{2.5}$ was the main factor in approximately 1.4 million deaths in China in 2015 caused by strokes ($PM_{2.5}$ is responsible for 40% of deaths from stroke), lung cancer (24%), and heart disease (27%) [4]. Therefore, accurate prediction of $PM_{2.5}$ not only prevents health problems but also gives the government early warning of environmental action.

Recently, many deep neural networks have been proposed to predict air pollution. The findings of their study demonstrate that deep neural networks perform better and can learn more deep features than conventional machine learning techniques [5, 6]. In light of this, we propose a fusion-based deep neural network to construct our predicted models: BiLSTM−BiGRU. The first layer of the deep neural network, *i.e.,* BiLSTM layer, extracts preliminary features, and the following three layers of the BiGRU model further extract the deep features. This fusion network achieves an accurate prediction result for the $PM_{2.5}$.

1.1. Related Works

Recent research has shown that recurring neural network (RNN), long short-term memory (LSTM), and convolutional neural network (CNN) models are very good at dealing with time series data.

Qin *et al.* [7] proposed a hybrid CNN−LSTM model to predict the urban $PM_{2.5}$ concentration. But this model has faced three key challenges [8]. First, CNN finds it challenging to extract the deep features from pollution data. Second, CNN-LSTM finds it challenging to extract spatio-temporal features from multiple pollutants. Third, LSTM finds it challenging to extract features from high-dimensional data. Tong *et al.* [9] used BiLSTM−RNN to predict the $PM_{2.5}$ in the southeast region of the USA. In this study, they analyze the spatiotemporal interpolation of different air pollutants. Finally, Liu *et al.* [10] applied the GRU model for the prediction of $PM_{2.5}$ in Beijing, China. In this research, they divided the geographical region into different groups and performed their analysis.

1.2. Contributions of the paper

The primary contributions of this paper can be summarized as follows:

- The fusion-based model, with BiLSTM as a base, is implemented to avoid vanishing gradient problems and to

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224.

Send correspondence to P. Dey: `d22124678@tudublin.ie`

explore the initial features of pollutants and meteorological data. BiGRU is adopted to further extract deep features and reduce the vanishing gradient problem.

- The FC layer is used to generate a precise prediction of $PM_{2.5}$ concentration. The experiment results demonstrate that our proposed model outperforms the other state-of-the-art models.
- In the spirit of reproducible research, we release the code to generate the results of this paper via `https://github.com/Prasanjit-Dey/PM2.5`.

2. PROPOSED METHOD

2.1. Framework overview

The training method for fusion neural networks involves mapping the original input to the target result. The air pollutants and the meteorological parameters from the previous r hours are fed into the fusion network: $x = \{x_t, \ldots, x_{t-i}, \ldots, x_{t-r-1}\}$, $x_{t-i} \in R^m$ (where m represents pollutants and meteorological parameters). The network predicted n hours of $PM_{2.5}$ concentration: $y = \{y_{t+1}, \ldots, y_{t+j}, \ldots, y_{t+n}\}$, $y_{t+j} \in R$, where y_{t+j} represents the prediction outcome. The framework is designed using a three-module architecture. In the first module, 50 units of the BiLSTM layer are extracted preliminary features from the training data. The second module, 50 units BiGRU further extracts the deep feature. The choice of the number of units in BiLSTM and BiGRU layers is selected based on various considerations, including the data's complexity and dimensionality, computational efficiency, and model performance. To regularize the training process, we also added dropout layers after each of the BiLSTM and BiGRU layers. The third module, the FC layer, received the output of the BiGRU layers and complete the $PM_{2.5}$ prediction as $y = \{y_{t+1}, \ldots, y_{t+j}, \ldots, y_{t+n}\}$. The framework is shown in Fig. 1.

2.2. BiLSTM−BiGRU

Initially, air pollutants data and meteorological parameters are fed into the both forward and backward direction of LSTM in the time-series order $x = \{x_t, \ldots, x_{t-i}, \ldots, x_{t-r+1}\}$ for preliminary features extraction. Then, each unit of forward and backward LSTM gates, namely the forget gate, the input gate, and the output gate, performs the feature extraction in the training data. Finally, forward and backward LSTM combine the features and generates the preliminary features in time series order $O = \{O_t, \ldots, O_{t-i}, \ldots, O_{t-r+1}\}$.

Therefore, BiLSTM used two hidden layers called forwards hidden layers ($h_t^{\xrightarrow{f}}$) and backward hidden layers ($h_t^{\xleftarrow{b}}$). Finally, the preliminary features are generated by combining $h_t^{\xrightarrow{f}}$ and $h_t^{\xleftarrow{b}}$. The following equation implements the BiLSTM network:

$$h_t^{\xrightarrow{f}} = tanh\left(W_h^{\xrightarrow{f}} x_t + W_h^{\xrightarrow{f}} h_{t-1}^{\xrightarrow{f}} + b_h^{\xrightarrow{f}}\right) \quad (1)$$

$$h_t^{\xleftarrow{b}} = tanh\left(W_h^{\xleftarrow{b}} x_t + W_h^{\xleftarrow{b}} h_{t-1}^{\xleftarrow{b}} + b_h^{\xleftarrow{b}}\right) \quad (2)$$

$$O_t = W^{\xrightarrow{f}} h_t^{\xrightarrow{f}} + W^{\xleftarrow{b}} h_t^{\xleftarrow{b}} + b_o \quad (3)$$

Where, O is denoted as output features, W is denoted as weight matrix, x is denoted as training data, and b is denoted as bias.

Next, output features $O = \{O_t, \ldots, O_{t-i}, \ldots, O_{t-r+1}\}$ were fed into forward and backward GRU for final feature extraction. Then, each unit of forward and backward GRU gates, namely the update gate and the reset gate, performs the feature extraction. Finally, the forward and backward GRU combine the features and generate the deep features in time series order $y = \{y_t, \ldots, y_{t-i}, \ldots, y_{t-r+1}\}$. The following equation implements the BiGRU network:

$$h_t^{\xrightarrow{f}} = tanh\left(W_h^{\xrightarrow{f}} O_t + W_h^{\xrightarrow{f}} h_{t-1}^{\xrightarrow{f}} + b_h^{\xrightarrow{f}}\right) \quad (4)$$

$$h_t^{\xleftarrow{b}} = tanh\left(W_h^{\xleftarrow{b}} O_t + W_h^{\xleftarrow{b}} h_{t-1}^{\xleftarrow{b}} + b_h^{\xleftarrow{b}}\right) \quad (5)$$

$$y_t = W^{\xrightarrow{f}} h_t^{\xrightarrow{f}} + W^{\xleftarrow{b}} h_t^{\xleftarrow{b}} + b_y \quad (6)$$

The extracted features are passed to the FC layers in the prediction state to predict the $PM_{2.5}$ concentration.

3. RESULTS & DISCUSSION

3.1. Dataset

The experiment used past pollution levels and meteorological parameters from Wanshou Xigong, China, between March 1, 2013, and February 28, 2017 (each hour interval total 35060 hours) [11]. The dataset contains 12 pollutant and meteorological variables, such as $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, O_3, temperature, air pressure, humidity, rain, wind direction, and wind speed. The models are trained on 26300 hours of data, and the remaining 8760 hours of data are utilized to test the models. we calculate the correlations between the data. Fig. 2 shows the correlations between the data and, it shows the highest correlation between $PM_{2.5}$ and other parameters. Therefore, in this study, $PM_{2.5}$ is chosen as the target variable.

3.2. Comparative analysis with state-of-the-art methods

Fig. 3 shows how different models can be generalized to the same test data. The x-axis of Fig. 3 has a length of 180 hours, meaning that the first 180 consecutive lead times were chosen from the test data for testing the performance of the prediction models. The red curve represents the actual test data, and the green curve represents the predicted value. Fig. 3 shows that the proposed model is more proficient at generalization than five state-of-the-art prediction models.

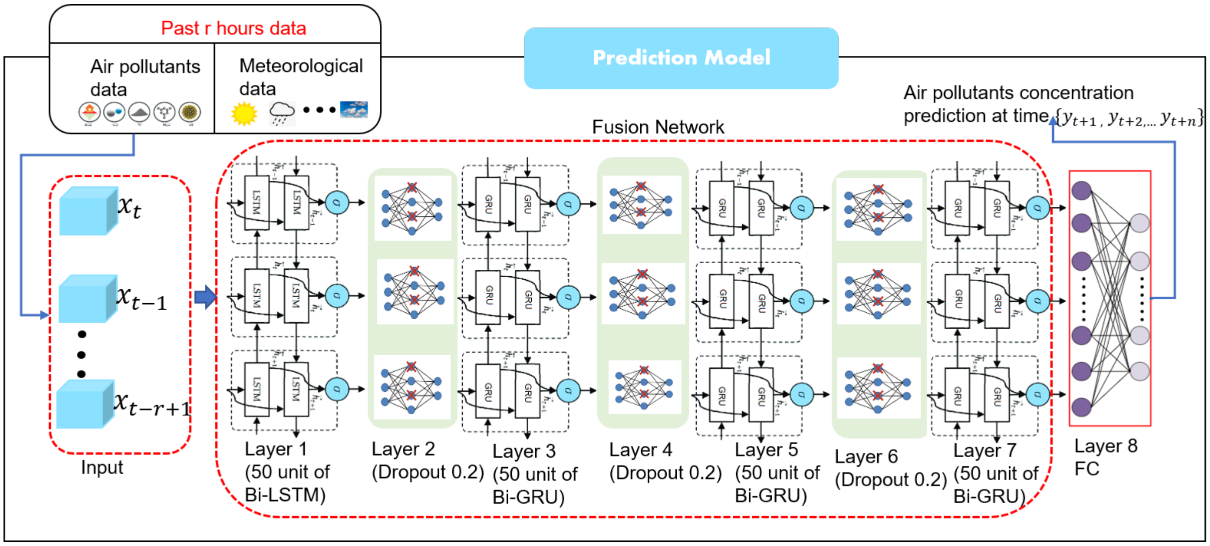

Fig. 1. Frameworks of the fusion-based deep neural network for $PM_{2.5}$ concentration prediction. x_{t-i} is the pollutants concentration and meteorological data input in the model for each timestamp.

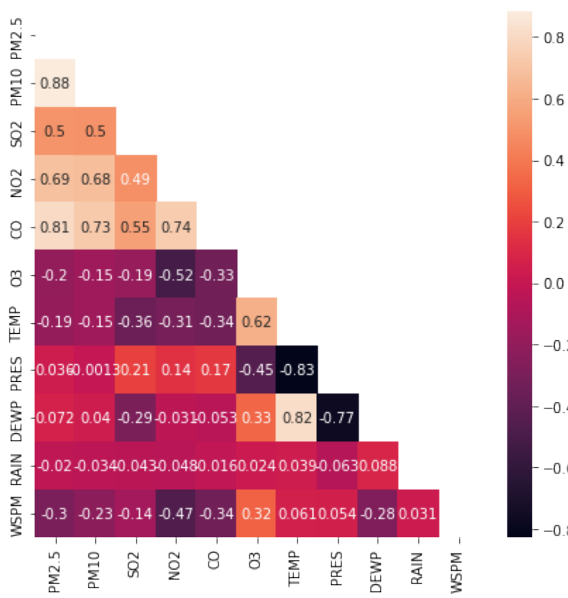

Fig. 2. Illustrated the correlation between the data. $PM_{2.5}$ has the highest correlation with other parameters.

To further measure the accuracy of the models, we used the first 180 hours of test data as input to predict $PM_{2.5}$ concentration. We compared the BiLSTM−BiGRU (proposed) model with the five state-of-the-art prediction models: CNN, LSTM, GRU, CNN−LSTM, and CNN−GRU. Table 1 shows the RMSE, MAE, and correlation (R^2) of the $PM_{2.5}$ concentration prediction values over different lead times. Table 1 shows that the proposed model has a better prediction result and can better handle the long-term prediction problem. The proposed model achieved values of RMSE, MAE, and R^2 of 22.0, 3.51, and 0.96, respectively, for 180 hours, which are at least 2.2 (RMSE) and 0.23 (MAE) lower values than those of other models. The data presented in Table 1 showed that the BiLSTM−BiGRU model is effective at predicting $PM_{2.5}$ concentrations.

	Methods	60h	90h	120h	150h	180h
RMSE	CNN	12.2	20.3	21.2	21.7	40.9
	CNN−LSTM	13.9	16.1	16.5	18.6	28.6
	CNN−GRU	22.6	23.3	22.8	21.1	30.6
	GRU	11.9	13.3	14.1	13.6	24.2
	LSTM	8.64	10.3	13.0	12.8	24.8
	Proposed	8.13	10.2	12.4	12.0	22.0
MAE	CNN	3.03	3.53	3.65	3.76	4.88
	CNN−LSTM	3.55	3.75	3.73	3.97	4.53
	CNN−GRU	4.62	4.65	4.52	4.29	4.79
	GRU	3.25	3.35	3.32	3.20	3.90
	LSTM	2.61	2.71	2.93	2.93	3.74
	Proposed	2.51	2.60	2.78	2.82	3.51
R^2	CNN	0.91	0.76	0.89	0.92	0.83
	CNN−LSTM	0.86	0.81	0.91	0.93	0.90
	CNN−GRU	0.72	0.68	0.85	0.92	0.88
	GRU	0.89	0.87	0.94	0.97	0.94
	LSTM	0.95	0.93	0.96	0.98	0.95
	Proposed	0.95	0.93	0.96	0.98	0.96

Table 1. Comparison of the performance of the proposed model with other models for the prediction of $PM_{2.5}$ concentration over multiple lead times.

4. CONCLUSION & FUTURE WORK

In this paper, we propose a fusion-based deep neural model called BiLSTM−BiGRU. The primary purpose of the model is to predict the $PM_{2.5}$ concentration in the specified city. The BiLSTM model is used to extract the preliminary features of the pollutants and meteorological data. The BiGRU is used

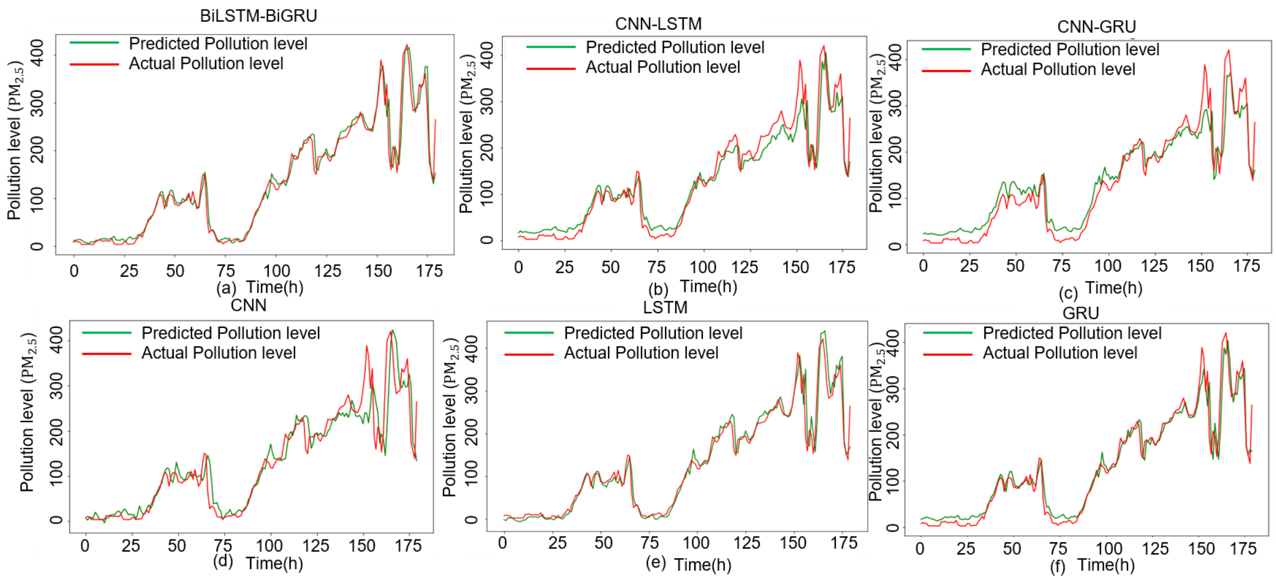

Fig. 3. Fitting trend of the different models for the 0 to 180 hours lead times. (a)−(f) represent the fitting trends of BiLSTM−BiGRU, CNN−LSTM, CNN−GRU, CNN, LSTM and GRU models.

to extract the deep features of the data output by the BiLSTM layer. The experimental results indicate that the BiLSTM-BiGRU model is more accurate for predictions, particularly in handling long-term lead time prediction. In the future, location data will be included among the input features used to make more accurate predictions.

5. REFERENCES

[1] G. Zou, B. Zhang, R. Yong, D. Qin, and Q. Zhao, "FDN-learning: Urban $PM_{2.5}$-concentration spatial correlation prediction model based on fusion deep neural network," *Big Data Research*, vol. 26, p. 100269, 2021.

[2] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2412–2424, 2019.

[3] N. Danesi, M. Jain, Y. H. Lee, and S. Dev, "Monitoring atmospheric pollutants from ground-based observations," in *Proc. IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*. IEEE, 2021, pp. 98–99.

[4] C. Song, J. He, L. Wu, T. Jin, X. Chen, R. Li, P. Ren, L. Zhang, and H. Mao, "Health burden attributable to ambient $PM_{2.5}$ in China," *Environmental pollution*, vol. 223, pp. 575–586, 2017.

[5] A. Dairi, F. Harrou, S. Khadraoui, and Y. Sun, "Integrated multiple directed attention-based deep learning for improved air pollution forecasting," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[6] P. Dey, S. Dev, and B. Schoen-Phelan, "NeSDeepNet: A fusion framework for multi-step forecasting of near-surface air pollutants," in *2023 Photonics & Electromagnetics Research Symposium (PIERS)*. IEEE, 2023.

[7] D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, and B. Zhang, "A novel combined prediction scheme based on CNN and LSTM for urban $PM_{2.5}$ concentration," *IEEE Access*, vol. 7, pp. 20 050–20 059, 2019.

[8] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

[9] W. Tong, L. Li, X. Zhou, A. Hamilton, and K. Zhang, "Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN," *Air Quality, Atmosphere & Health*, vol. 12, no. 4, pp. 411–423, 2019.

[10] B. Liu, S. Yan, J. Li, Y. Li, J. Lang, and G. Qu, "A spatiotemporal recurrent neural network for prediction of atmospheric $PM_{2.5}$: a case study of Beijing," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 3, pp. 578–588, 2021.

[11] S. Zhang, B. Guo, A. Dong, J. He, Z. Xu, and S. X. Chen, "Cautionary tales on air-quality improvement in Beijing," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2205, p. 20170457, 2017.