

Application of Informer for time-series NO2 prediction

Hye Yeon Sin*, Minchul Kang**, Joonsung Kang***

*Associate professor, Department of Pharmacy, College of Pharmacy, Duksung Women's University, Seoul, Korea

**Assistance professor, Graduate School of Consulting and Clinical Psychology, CHA University, Gyeonggi-do, Korea

***Professor, Department of data science, Gangneung-Wonju National Univeristy, Gangneung, Korea

[Abstract]

In this paper, we evaluate deep learning time series forecasting models. Recent studies show that those models perform better than the traditional prediction model such as ARIMA. Among them, recurrent neural networks to store previous information in the hidden layer are one of the prediction models. In order to solve the gradient vanishing problem in the network, LSTM is used with small memory inside the recurrent neural network along with BI-LSTM in which the hidden layer is added in the reverse direction of the data flow.

In this paper, we compared the performance of Informer by comparing with other models (LSTM, BI-LSTM, and Transformer) for real Nitrogen dioxide (NO₂) data. In order to evaluate the accuracy of each method, mean square root error and mean absolute error between the real value and the predicted value were obtained. Consequently, Informer has improved prediction accuracy compared with other methods.

▶ **Key words:** Deep learning, Prediction, LSTM, BI-LSTM, Transformer, Informer

[요 약]

본 논문에서는 딥러닝 시계열 예측 모형을 평가한다. 최근 연구에 따르면 이 모형은 ARIMA와 같은 기존 예측 모형보다 성능이 우수하다고 결론짓는다. 그 중 히든 레이어에 이전 정보를 저장하는 순환 신경망이 이를 위한 예측 모형 중 하나이다. 네트워크의 그래디언트 소실 문제를 해결하기 위해 LSTM은 데이터 흐름의 반대 방향으로 숨겨진 레이어가 추가되는 BI-LSTM과 함께 순환 신경망 내부의 작은 메모리로 사용된다. 본 논문은 서울의 2018년 1월 1일부터 2022년도 1월 1일까지의 NO₂ 자료에 대해 Informer의 성능을 LSTM, BI-LSTM, Transformer와 비교하였다. 이에 실제 값과 예측값 사이의 평균 제곱근 오차와 평균 절대 오차를 구하였다. 그 결과 Test 데이터 (2021.09.01.~2022.01.01.)에 대해 Informer는 다른 방법에 비해 가장 높은 예측 정확도 (가장 낮은 예측 오차: 평균 제곱근 오차: 0.0167, 평균 절대 오차: 0.0138)를 보여 타 방법에 비해 그 우수성을 입증하였다. Informer는 당초 취지와 부합되게 다른 방법들이 갖고 있는 장기 시계열 예측에 있어서의 문제점을 개선하는 결과를 나타내고 있다.

▶ **주제어:** 딥러닝, 예측, 장단기 메모리, 양방향 장단기 메모리, 변환기, 정보 제공자

-
- First Author: Hye Yeon Sin, Co-author: Minchul Kang, Corresponding Author: Joonsung Kang
 - *Hye Yeon Sin (hyshin@duksung.ac.kr), Department of Pharmacy, College of Pharmacy, Duksung Women's University
 - **Minchul Kang (kkmmcc1122@naver.com), Graduate School of Consulting and Clinical Psychology, CHA University
 - ***Joonsung Kang (mkang@gwnu.ac.kr), Department of data science, Gangneung-Wonju National Univeristy
 - Received: 2023. 06. 08, Revised: 2023. 07. 17, Accepted: 2023. 07. 20.

I. Introduction

Time series data prediction has been studied for a long time in various fields such as finance, climate, medicine, and transportation. Traditionally, there are statistics-based forecasting models such as ARIMA and VAR. Recently, deep learning models that effectively learn the nonlinear and irregular characteristics of time series data are being actively researched. More specifically, deep learning methods offer a lot of promise for time series forecasting, such as the automatic learning of temporal dependence and the automatic handling of temporal structures like trends and seasonality. According to such experimental results, it has been shown that deep learning models perform better than the traditional ARIMA or SARIMA methodology [1-5]. Notably, with the explosive increase of various deep learning models, the best model among them needs to be taken into account for better accuracy and efficiency.

The main purpose of this paper is to propose Informer as the best deep learning based time series prediction model for Seoul's air pollution (NO₂) prediction in comparison with LSTM (Long-Short Term Memory), BI-LSTM (Bidirectional LSTM), and Transformer [6-8].

II. Description of 4 deep learning prediction models

Deep learning is an artificial intelligence (AI) approach that teaches computers to process data in ways inspired by the human brain. The models can recognize complex patterns in pictures, text, sound and other data to generate accurate insights and predictions. Generally, a deep learning's component, an artificial neural network consists of input layer, hidden layers, and output layers. In this paper, among various kinds of neural networks, we will use the concepts of RNN and its applications shown below in order to analyze our

data. For our data analysis, Python deep learning package Pytorch (<https://pytorch.org/>) is used.

1. Recurrent Neural Networks (RNNs)

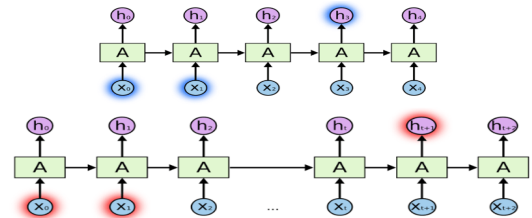


Fig. 1. RNN structure.

RNN is a type of artificial neural network that uses sequential or time series data (Fig. 1.). This deep learning algorithm is commonly used for sequence or timing problems such as language conversion, natural language processing (NLP), speech recognition, image captions, time series data forecasting, our analysis. The main problems of conventional RNN, however, is in long-term dependency problems. Therefore, LSTM was introduced for processing such sequence data, and thus solving the vanishing gradient problem that RNNs have [9-10].

2. LSTM

LONG SHORT-TERM MEMORY NEURAL NETWORKS

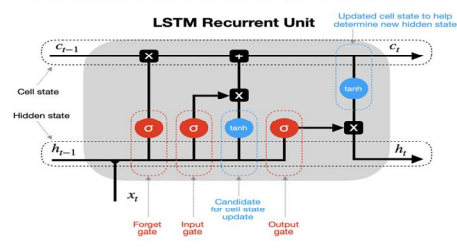


Fig. 2. LSTM.

LSTM has been used to better predict NO₂. Previous study showed that the LSTM model has high application value in NO₂ concentration prediction [11-14]. Suhartono (2019) proposed a hybrid model combining time series regression (TSR) and LSTM with a higher accuracy to forecast

NO2 in Surabaya City, Indonesia [15]. Drewil and Bahadili (2022) forecasted air pollution using LSTM method. In this sense, LSTM has become a major prediction model in deep learning [16]. In Fig. 2., LSTM adjusts the information to be contained in the cell state through a total of three gates and transfers it to the next state [9, 17]. The forget gate is a gate that determines whether to discard or use information. By using the sigmoid function, we have a number between 0 and 1. By multiplying the number derived here with c_{t-1} of the cell state, we decide whether to use the value of the previous state or not. The input gate determines which of the new information to store in the cell state. In the tanh layer, a new information candidate vector is determined, and in the input layer, the sigmoid function is used to determine which information among the candidates to use. After that, the value deleted or used by the forget gate and the new information determined through the input gate are added. Finally, in the output gate, a sigmoid is applied to the input value to determine which value to derive as the output value from the cell state. Then, the value of the cell state is multiplied by the value obtained by tanh, and only the required output value is produced.

LSTM can solve time series data with longer sequences than existing RNNs [9, 11-12, 18].

3. BI-LSTM

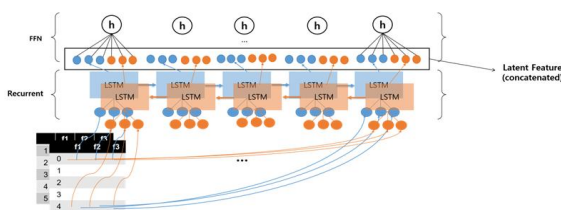


Fig. 3. BI-LSTM many to many structure.

Wu et al. (2023) show that the hybrid model, the Res-GCN-BiLSTM model was better adapted and improved the prediction accuracy, with nearly 11% improvements in mean absolute error for NO2 compared to the best performing baseline model.

[19]. Verma et al. (2018) showed that the predictions can be significantly improved using BI-LSTM that model the long-term, short-term and immediate effects of PM2.5 severity levels [19]. BI-LSTM becomes a better prediction model mechanism for NO2 than LSTM. BI-LSTM is a time-stepped bidirectional long-latency RNN on time-series extended data. This can be useful if you want such a standard RNN to train on the entire time series each time. In Fig. 3., BI-LSTM adds an LSTM layer that processes backward (1) to the existing LSTM layer. The final hidden state outputs a vector connecting the hidden states of the two LSTM layers. In addition to concatenation, adding or averaging can be applied in various ways. There is a limitation that the result shows a tendency to converge based on the previous pattern in RNN or LSTM. Both this convergence problem and long sequence one lead to introduce BI-LSTM instead of BI-RNN or LSTM [10, 13, 20].

4. Transformer

Hickman et al. (2022) predicted European ozone air pollution using Transformer [21]. It has recently taken more interest among various deep learning models.

A major issue in long sequence time series prediction (LSTF) is to increase the prediction capacity in order to satisfy the demand for increasingly longer sequences. In this respect, (a) the ability to classify over a long ranges (b) efficient operation of long sequence inputs and outputs is required. Accordingly, the Transformer model (Fig. 4.) showed better performance than RNN in capturing long-range dependencies [22]. The self-attention mechanism reduces the maximum length of the network signal propagation path to the theoretically shortest $O(1)$ and avoids repetitive structures. Consequently, it showed excellent potential in LSTF problems. Nevertheless, this mechanism violates requirement (b) because it consumes L -order computation and memory for inputs and outputs of length L . Thus, the

Transformer structure is not applied to LSTF problems well.

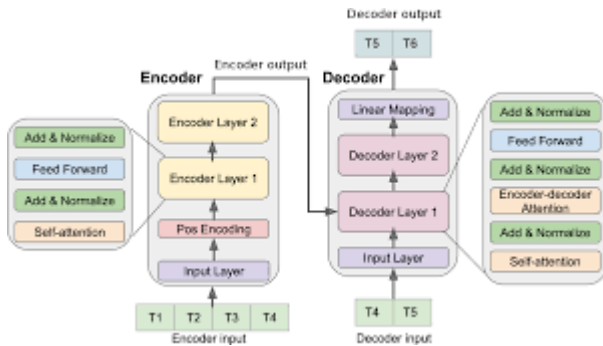


Fig. 4. Transformer.

5. Informer

As mentioned earlier, the Transformer model has problems in direct application to LSTF due to time complexity, high memory usage, and inherent limitations of the encoder-decoder architecture. For those problems, an efficient LSTF converter called Informer was recently proposed [23]. This solves the LSTF problem with the encoder-decoder structure. An outline of the model is shown in the following figure [23].

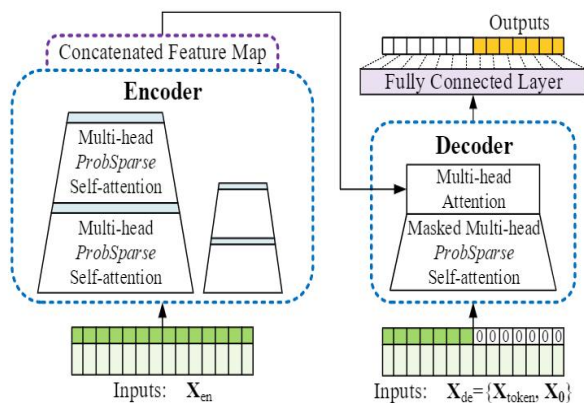


Fig. 5. Informer.

Informer utilizes the ProbSparse self-attention mechanism when the encoder receives a long sequence (Fig. 5.). The small size cloned encoder next to it increases the robustness of the model. When the decoder inputs a long sequence, the predicted object portion should be padded with zeros. The decoder predicts the output once by

performing the concatenated feature map generated by the encoder and encoder-decoder attention. In previous studies, predictions were made with RNN models, LSTMs, Transformer, and the Informer model which showed the highest prediction accuracy [24-25]. The Informer model greatly improves LSTF problem solving owing to its structural advantages. For more details on the Informer model, see Zhou et al. (2012) [23].

III. Data preparation

1. Data acquisition

Information on daily average air pollution in Seoul provides daily information on average air pollution such as air quality index, fine dust, ozone, nitrogen dioxide, carbon monoxide, and sulfur dioxide.

Among them, daily air pollution (NO₂) data (Seocho-gu) in Seoul from January 1, 2018 to January 1, 2022 were selected as a sample (<https://data.seoul.go.kr/dataList/OA-2218/S/1/datasetView.do>).

2. Preprocessing

We first preprocessed the downloaded NO₂ data by replacing the missing 19 values by the mean. The criterion for the outlier was recognized as a value outside the 1.5-fold interquartile range. Figure 6 shows training and test data for daily average NO₂ data. Since the mean and variance in this plot are largely constant, it can be considered a stationary time series. Therefore, no other transformations were required. MinMax scaling was performed on the training data.

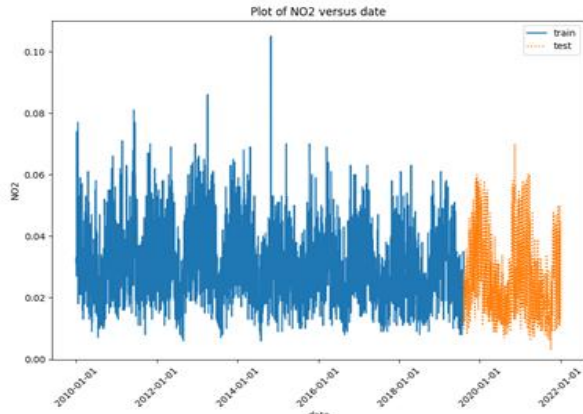


Fig. 6. Plot of daily average NO2 versus date for train (including validation data) and test data.

3. Train data, validation data and test data.

We analyzed 877 NO2 observations out of the 4383 total NO2 data as test data (20%). The number of NO2 observations for train data is 2804. The 702 observations is for validation. The proportion of train, validation and test data is 0.64:0.16:0.2.

IV. Experiment and Results

1. Computing environment

This model utilized Google colab with the Python package Pytorch.

2. Settings

For comparative analysis, LSTM, BI-LSTM, Transformer, and Informer are trained. Prediction accuracy was evaluated by comparing predicted values with out-of-sample predictions with actual values in the test data.

3. Results

The train data was initially used to train LSTM and BI-LSTM, Transformers, and Informers. At the same time, it is trained using validation data to ensure that the error is small enough. The prediction was made by setting the window size to 100, and the optimal model was derived by adjusting the number of epochs, the loss function, the number of input layers, and the number of

hidden layers, and so forth. As a result, the Adam optimizer using 0.001 as the learning rate and the Huber loss function (delta=0.1) were utilized. LSTM and BI-LSTM were prematurely stopped at the 68th and 70th epochs, respectively. For BI-LSTM, dropout was set to 0.5 and batch normalization was performed with momentum of 0.5.

These can be summarized in Tables 1 and 2, respectively.

Table 1. LSTM network

Layer (type)	Output Shape	Param #
Istm_1 (LSTM)	(None, 1, 100)	80400
batch_normalization	(None, 1, 100)	400
Istm_1 (LSTM)	(None, 100)	80400
batch_normalization	(None, 1, 100)	400
dense (Dense)	(None, 1)	101
Total params: 161,701 Trainable params: 161,301 Non-trainable params: 400		

Table 2. BI-LSTM network

Layer (type)	Output Shape	Param #
bidirectional_1 (Bidirectional)	(None, 1, 100)	60400
bidirectional_1 (Bidirectional)	(None, 1, 100)	60400
batch_normalization_4 (Batch Normalization)	(None, 100)	400
dense_6 (Dense)	(None, 1)	101
Total params: 121,301 Trainable params: 121,101 Non-trainable params: 200		

In this study, RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) were used to evaluate the prediction accuracy. In Fig. 7, BI-LSTM has lower RMSE than LSTM. Similarly, in Figure 8, the MAE of BI-LSTM was lower than that of LSTM. RMSE and MAE gradually declined over epochs.

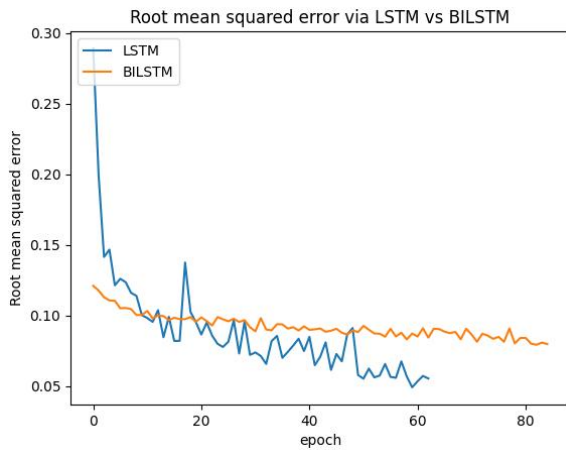


Fig. 7. RMSE

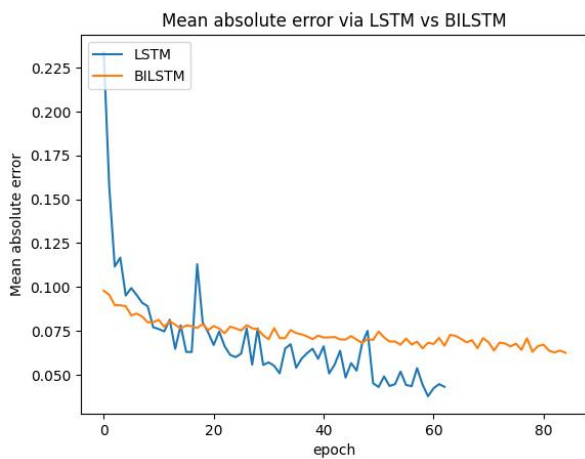


Fig. 8. MAE

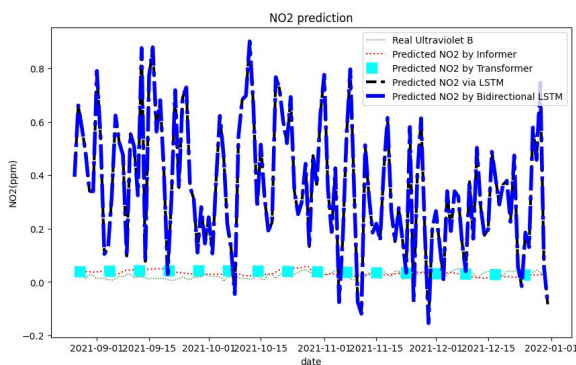


Fig. 9. Predicted values of NO2 by LSTM, BI-LSTM, Transformer, and Informer versus real NO2.

Table 3. Prediction error

	MAE	RMSE
LSTM	0.3016	0.2413
BI-LSTM	0.1639	0.2127
Transformer	0.0160	0.0187
Informer	0.0138	0.0167

Table 3 tells us that Informer is preferable over LSTM, BI-LSTM, and Transformer in terms of relatively smaller MAE and RMSE for test data. Informer and Transformer has 0.0138 and 0.0160 for MAE and 0.0167 and 0.187 for RMSE, respectively, concluding that Informer is a little bit better mechanism than Transformer. That's because Informer can deal better with the LSTF problem for a range of daily NO2 data.

Fig. 9 compared the predicted NO2 values by LSTM and BI-LSTM, Transformer, and Informer with the real NO2 value in the test set. For the specific period between September 1, 2021 and January 1, 2022 regarding test data, Informer and Transformer appeared to be much more accurate than LSTM and BI-LSTM, though they might be known as accurate prediction models. Informer is effective in that for predicting air pollution (NO2) in Seoul, solving an LSTF problem.

V. Comparison with other studies for NO2 concentration prediction models

So far, a few of prediction models for NO2, as the main air pollutant, has been introduced using deep learning methodology. Yammahi and Aung (2023) predicted NO2 concentration by ARIMA, SARIMA, LSTM, and nonlinear autoregressive neural network (NAR-NN) with both open- and closed-loop architectures, showing that predictions based on the open loop are better than those based on the closed loop [7]. In another study conducted by Heydari et al. (2021), a hybrid model based on long short-term memory (LSTM) and multi-verse optimization algorithm (MVO) was developed to predict NO2 [8]. In the paper by Liu et al. (2021), a daily NO2 concentration prediction model in Beijing based on LSTM is constructed and they showed that the LSTM model has high application value in NO2 concentration prediction [13]. Wu et al. (2023) proposed a hybrid model called Res-GCN-BiLSTM combining the residual neural network (ResNet),

graph convolutional network (GCN), and BI-LSTM, for predicting short-term NO₂, not long-term behavior of it. They utilized a hybrid model combining LSTM-based methods and other ones but they did not include the recently presented methods such as Transformer-based models. In this respect, our prediction models, Informer for NO₂ is meaningful in meteorology perspective. This research focused on Seoul's NO₂ data only but there are a few of sources of air pollutants to be studied in the future. Moreover, a hybrid of deep learning methodologies or other newly presented method needs to be assessed for more exact air pollutant prediction.

VI. Conclusions

In the paper, LSTM, BI-LSTM, Transformer, and Informer were applied to NO₂ with the performance assessment. As a result, Informer showed the best prediction accuracy for Seoul's air pollution (NO₂) prediction, followed by Transformer, BI-LSTM and LSTM. Our NO₂ data is a typical example of LSTF problems. The results show that among 4 deep learning methods, Informer can effectively process those problems. Another manuscript titled 'The prediction model of Ultraviolet-B (UV-B) using Deep learning model (Bidirectional LSTM) in comparison with LSTM and SARIMA' has analyzed Ultraviolet-B data by comparing LSTM, BI-LSTM, and SARIMA [25]. This data showed strong seasonality. When fitting such a data, deep learning models such as Transformer-based models or LSTM (or BI-LSTM) does not quite well decompose time series into seasonality and trend-cycle components, causing interpretation problems. It was expected that SARIMA performs better than other models because of this seasonality but it did not in reality. It's, however, enough to mention that based on results from a few of other previous studies, SARIMA did better for short Term prediction only but LSTM or BI-LSTM would definitely do better [26]. In fact, for UV-B data, Transformer or Informer did not even

converge completely with slower speed. It does not express seasonality well and still causes a memory bottleneck.

REFERENCES

- [1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, "Time Series Analysis, 4th Edition", Wiley Series in Probability and Statistics, June 2008. DOI:10.1002/9781118619193
- [2] C. Lee and J. Kim, "A study on the short-term prediction of power consumption by using the ARIMA model", Journal of the Korean Data Analysis Society, Vol. 19, No. 3, pp. 1349- 1362, June 2017. DOI:10.37727/jkdas.2017.19.3.1349
- [3] D.-C. Han, D. W. Lee, and D. Y. Jung, "A Study of the Traffic Volume Correction and Prediction Using SARIMA Algorithm", J. Korea Inst. Intell. Transp. Syst., Vol. 20, No.6,pp.1-13,2021, <https://doi.org/10.12815/kits.2021.20.6.1>
- [4] J. Korstanje, "The SARIMAX Model", Advanced Forecasting with Python, Springer, pp 125-131, July 2021.
- [5] S. S.-Namini, N. Tavakoli, A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series", 2018 17th IEEE International Conference on Machine Learning and Applications, pp. 1-8, 2018. DOI: 10.1109/ICMLA.2018.00227
- [6] P. T. Yamak, L. Yujian, P. K. Gadosey, "A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting", ACAI 2019: Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence, pp.49-55. December 2019, <https://doi.org/10.1145/3377713.3377722>
- [7] A. A. Yammahi and Z. Aung, "Forecasting the concentration of NO₂ using statistical and machine learning methods: A case study in the UAE", Heliyon, Vol. 9, Issue 2, February 2023, e12584
- [8] A. Heydari, M. M. Nezhad, D. A. Garcia, F. Keynia, and L. D. Santoli, "Air pollution forecasting application based on deep learning model and optimization algorithm", Clean Technologies and Environmental Policy, 2021.
- [9] J. Ilonen, J.-K. Kamarainen, and J. Lampinen, "Differential evolution training algorithm for feed-forward neural networks", Neural Processing Letters, Vol. 17, No. 1, pp. 93-105, 2003, <https://doi.org/10.1023/A:1022995128597>
- [10] J. Kim and J.-Y. Kim, "Comparative analysis of performance of BI-LSTM and GRU algorithm for predicting the number of Covid-19 confirmed cases", Journal of the Korea Institute of Information and Communication Engineering, Vol. 26, No. 2, pp.187-192, 2022.
- [11] S. Hochreiter, J. Schmidhuber, "Long short-term memory", Neural Computation, Vol. 9, No. 8, pp. 1735-1780,1997. <https://doi.org/10.1162/neco.1997.9.8.1735>

- [12] J. Jung and J. Kim, "A Performance Analysis by Adjusting Learning Methods in Stock Price Prediction Model Using LSTM", *Journal of Digital Convergence*, Vol. 18, No. 11, pp.259-266,2020, <https://doi.org/10.14400/JDC.2020.18.11.259>
- [13] B. Liu, X. Yu, Q. Wang, S. Zhao, and L. Zhang, "A Long Short-Term Memory Neural Network for Daily NO₂ Concentration Forecasting", *International Journal of Information Technology and Web Engineering (IJITWE)*, Vol. 16, Issue 4, 2021.
- [14] C-L. Wu, H-D. He, R-F. Song, X-H. Zhu, Z-R. Peng, Q.-Y. Fu, and J. Pan, "A hybrid deep learning model for regional O₃ and NO₂ concentrations prediction based on spatiotemporal dependencies in air quality monitoring network", *Environmental Pollution*, Vol. 320, 1, March 2023.
- [15] Suhartono, H. Prabowo and S-F. Fam, "A Hybrid TSR and LSTM for Forecasting NO₂ and SO₂ in Surabaya", *Home Soft Computing in Data Science Conference paper*, First Online: 24 September 2019.
- [16] G. I. Drewil, R. J. A-Bahadili, "Air pollution prediction using LSTM deep learning and metaheuristics algorithms", *Measurement: Sensors*, Vol. 24, 2022, 100546
- [17] L. Nashold and R. Krishnan, "Using LSTM and SARIMA Models to Forecast Cluster CPU Usage", *arXiv:2007.08092[cs.LG]*, pp. 1-11, Jul 2020, <https://doi.org/10.48550/arXiv.2007.08092>
- [18] Preeti, R. Bala, R. P. Singh, "Financial and Non-Stationary Time Series Forecasting using LSTM Recurrent Neural Network for Short and Long Horizon", *International Conference on Computing and Networking Technology (ICCNT)*, pp.1 - 7, Kanpur, India, July 2019, DOI:<https://doi.org/10.1109/ICCNT45670.2019.8944624>
- [19] I. Verma, R. Ahuja, H. Meisheri, and L. Dey, "Air Pollutant Severity Prediction Using Bi-Directional LSTM Network", 2018 *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, December 2018, DOI:10.1109/WI.2018.00-19.
- [20] S. Ko, H.-Y. Yun, and D.-M. Shin, "Electronic Demand Data Prediction using Bidirectional Long Short Term Memory Networks", *Journal of Software Assessment and Valuation*, Vol. 14, No. 1, pp. 33-40, 2018.
- [21] S. H. M. Hickman, P. T. Griffiths, P. Nowack, E. Alhajjar, and A. T. Archibald, "Forecasting European Ozone Air Pollution With Transformers", *Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022*.
- [22] A. Zeng, M. Chen, L. Zhang, Q. Xu, "Are Transformers Effective for Time Series Forecasting?", *arXiv:2205.13504v3 [cs.AI]*, pp. 1-15, 2022.
- [23] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond Efficient Transformer for long Sequence Time-Series Forecasting", *arXiv:2012.07436v3 [cs.LG]*, pp 1-15, Mar 2021.
- [24] H. Wei, W.-S. Wang, X.-X. Kao, "A novel approach to ultra-short-term wind power prediction based on feature engineering and Informer", *Energy Reports*, Vol. 9, pp. 1236-1250, Dec 2023.
- [25] M. Kang and J. Kang, "The prediction model of Ultraviolet-B using Deep learning model (Bidirectional LSTM) in comparison with LSTM and SARIMA", *Journal of The Korea Society of Computer and Information (JKSCI)*, submitted, 2023.
- [26] T. Falatouri, F. Darbanian, P. Brandtner and C. Udokwu, "Predictive Analytics for Demand Forecasting - A Comparison of SARIMA and LSTM in Retail SCM", *Procedia Computer Science*, Vol. 200, pp. 993-1003, 2022.

Authors



Hye Yeon Sin received the B.S., in College of Pharmacy from, Duksung Women's University, Seoul, Korea. in 1984 and Pharm.D. degree in College of Pharmacy from Western University of Health Sciences

in Southern California, U.S.A., in 2006, respectively. Dr. Sin joined the faculty of the Department of Pharmacy, College of Pharmacy at Duksung Women's University, Seoul, Korea, in 2013. She is currently a Associate Professor in the Department of Pharmacy, College of Pharmacy at Duksung Women's University, Seoul, Korea. She is interested in parallel computing, internet and mobile computing, and cloud computing.



Minchul Kang received his B.S degrees in Statistics from Korea University, M.S. degrees in Art Therapy from Newyork University and Ph.D. degrees in Educational Counseling from Seoul National University. Dr. Kang joined

the faculty of Graduate School of Clinical and Counseling Psychology at CHA University, Gyeonggi-do Province, Korea, 2016.



Joonsung Kang received the B.S in Statistics from Korea university, Republic of Korea in 2002 and Ph.D. degrees in Biostatistics from University of North Carolina at Chapel Hill, USA, in 2009.

He is currently a Professor in the Department of Data Science, Gangneung-Wonju National university. He is interested in causal inference, deep learning and big data analysis.