International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparison of Data Augmentation Techniques for pm2.5 Time Series Prediction

Anibal Flores*[1], Hugo Tito-Chura[2], Deymor Centty-Villafuerte[3], Alejandro Ecos-Espino[4]

Submitted: 29/01/2024 **Revised**: 07/03/2024 **Accepted**: 15/03/2024

Abstract: Pm2.5 time series prediction is a very important task; in the literature, many approaches have been implemented, but just two use data augmentation to improve results. In this work, three data augmentation techniques are implemented and analyzed, two of them typical of the state of the art, thus the first (DA1) is time-warping + jittering, the second (DA2) is linear interpolation, and the third DA3 is polynomial interpolation which is proposed in this work. The performance of the data augmentation techniques is evaluated through four deep learning techniques including Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and Bidirectional GRU (BiGRU). In terms of RMSE, MAPE, and R^2, the results show that DA2 and DA3 are superior to DA1, between 3.62% to 4.04% respectively. While DA2 and DA3 present similar performances, the main difference between them is the higher computational cost of DA3 concerning DA2.

Keywords: *pm2.5 prediction; data augmentation; time series; deep learning*

1. Introduction

Air pollution is the contamination of the environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere [1]. One type of air pollutant is particulate matter 2.5 (pm2.5), which are very small airborne particles with a diameter of around 2.5 micrometers, which is less than the thickness of human hair. Many studies show that pm2.5 can induce a variety of chronic diseases [2], such as respiratory system damage [3], cardiovascular dysfunction [4], lung damage [5], and diabetes mellitus [6] among others.

Therefore, it is important to implement PM2.5 prediction models that allow estimating the future values of this pollutant in order to support the timely decision-making [7] of the corresponding responsible entities.

On the other hand, data augmentation is a strategy to increase the number of training instances from the existing ones to overcome underfitting [8],[9] and overfitting[10],[11] problems due to scarcity of data. It arose in the field of computer vision through techniques such as flipping, zooming, scaling, cropping, and rotation among others. Then it spread to other fields such as natural processing language (NLP) until it reached time series. In time series was initially applied to time series classification and later to time series regression.

In this study data augmentation techniques for time series regression are compared, thus three data augmentation techniques are implemented. The first DA1 is based on time-warping and jittering from the works [7] and [8] for short-term and solar radiation time series prediction. The second DA2 is based on linear interpolation implemented in the work [9] for solar radiation forecasting. Finally, the third DA3 has not been used yet, so this technique is proposed in this work, which consists of polynomial interpolation.

The Deep learning architectures that are implemented in this work are similar to those implemented in other related works. Thus, in this work, four deep learning models are implemented including LSTM, Bidirectional LSTM (BiLSTM), GRU, and Bidirectional GRU (BiGRU), for which pm2.5 time series of an environmental monitoring station located in the city of Ilo in southern Peru is used.

The main contributions of this work are summarized below:

- The proposal of a new data augmentation technique for pm2.5 time series forecasting based on polynomial interpolation.

- A comparative analysis of data augmentation techniques for pm2.5 time series forecasting including time-warping + jittering, linear interpolation, and polynomial interpolation.

- A comparative analysis of the performance of deep learning models with and without the use of the data augmentation techniques implemented for pm2.5 time series forecasting.

[1] *Universidad Nacional de Moquegua, Dpto. Ingeniería de Sistemas e Informática, Urb. Ciudad Jardin-Pacocha-Ilo, Moquegua 18611, PERU
ORCID ID : 0000-0003-4638-8242*
[2] *Universidad Nacional de Moquegua, Dpto. Ingeniería de Sistemas e Informática, Urb. Ciudad Jardin-Pacocha-Ilo, Moquegua 18611, PERU
ORCID ID : 0000-0001-8269-0655*
[3] *Universidad Nacional de Moquegua, Dpto. de Ciencias Sociales y Humanidades, Prolongación Calle Ancash S/N, Moquegua 18001, PERU*
[4] *Universidad Nacional de Moquegua, IINDEP, Calle Ancash S/N, Moquegua 18001, PERU
ORCID ID : 0000-0002-3452-5388*
** Corresponding Author Email: afloresg@unam.edu.pe*

The rest of the paper has been organized in a Related Work section where the pm2.5 time series regression works are described; then a Methodology section, where the activities in detail to implement data augmentation and deep learning models are described; next Results a Discussion section, where achieved results are described and discussed; and finally, a Conclusion section that concludes de research work.

2. Related Work

This section briefly describes chronologically related work to pm2.5 time series prediction.

Most of the related works have used LSTM, among them [12]–[18] and [18] some of them include decomposition techniques, and others are implemented combined with other techniques. In [12] and [14] Bidirectional LSTM (BiLSTM) and LSTM are used respectively. In [15] the authors added a decomposition technique named Singular Spectrum Analysis (SSA). In [16] LSTM is hybridized with Convolutional Neural Networks (CNN). In [19] LSTM is hybridized with a Genetic Algorithm (GA) and an Encoder-Decoder (ED) model. In [17] a Bidirectional LSTM model combined with a CNN model are used. And in [18], it is combined with decomposition techniques such as CEEMDAN and FCM. Similar to LSTM, GRU has also been used in [20] and [21], in the first case with data augmentation and in the second in a hybrid way with Q-Learning.

Support Vector Regression (SVR) and Random Forest have been used in works such as [22], [23], and [24]. The first uses SVR in a hybrid way with Quantum Particle Swarm Optimization (QPSO), the second adds a decomposition technique named hybrid modified variational mode decomposition and the last one uses just Random Forest.

Other related works such as [25]–[29] propose different models to those cited above. Thus in [25] is proposed a Hammerstein Recurrent Neural Networks (CHRNN), in [26] is proposed a Multiple Model Adaptive Unscented Kalman Filter (MMAUKF), in [24] a decomposition ensemble learning based on variation mode decomposition (VMD) and whale-optimization algorithm (IWOA) is proposed, in [28] an attention-based deep neural network is used and in [29] a deep belief network using PM2.5 and temperature data is proposed.

In [20] data augmentation with LSTM and GRU for pm2.5 time series prediction are used and in [30] data augmentation and Resnet. These are the only works that use the generation of synthetic data to improve the performance of regression models. In the first case, linear interpolation is used, and, in the second case, random over-sampling (ROS).

The results of related works are compared with the results of implemented models based on data augmentation and deep learning in this work's Results and Discussion section.

3. Methodology

This section describes in detail the developed methodology.

3.1 Data Preparation

The selected time series corresponds to the Pacocha station in Ilo City, located in the district of Pacocha, Province of Ilo, Moquegua region in southern Peru.

The data range from 2022-06-03 15:00 to 2023-04-17 22:00 and correspond to 7640 hours. The training data corresponds to 80% (6112 hours) and the remaining 20% (1528 hours) for testing. Fig. 1 shows training and test data graphically.

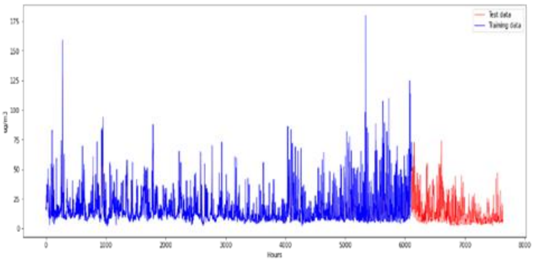

Fig. 1. Training and test data

3.2 Imputation

For time series imputation, there are several options, but due to the missing values (NA) correspond to short-gaps, and gaps between 1 to 4 NA values, imputeTS [31][32] library was used. ImputeTS is an R language library that implements several techniques including interpolations (spline, stinneman) and several based on moving averages (SMA, LWMA, EWMA, ARIMA), from these, Single Moving Average (SMA) with k=2 was selected. In total 48 missing values were filled.

3.3 Data Augmentation

In this phase, three data augmentation techniques are implemented: time-warping + jittering [33] (DA1), linear interpolation [20] (DA2), and polynomial interpolation (DA3) which is proposed in this work. Fig. 2 shows graphically these techniques with 5 synthetic features for the first couple of values in a time series.

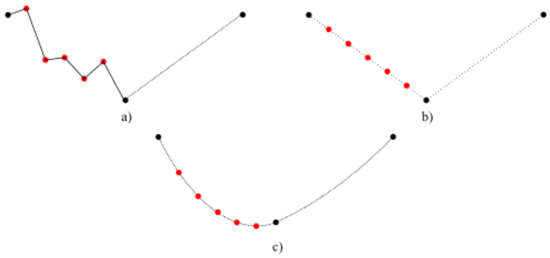

Fig. 2. Data augmentation techniques for time series regression: a) DA1 b) DA2 and c) DA3

3.3.1 Time-warping + jittering (DA1)

This data augmentation technique for time series regression is a mix between linear interpolation and random values, some values are linear interpolated and the random ones are generated considering a range of values. The random values allow the generation of an irregular curve, enriching the information of the time series. Fig. 1 a) graphically shows this technique.

In this work, the following parameters are used for this technique:

Block size	:	10
Subblock size	:	2
Precision	:	2
Window size	:	2

From an original time series of 6112 items, these parameters produce 61110 synthetic items.

3.3.2 Linear interpolation (DA2)

Linear interpolation was used in DA1 mixed with random values. However, DA2 works just with linear interpolation.

Given two points (*x1*,*y1* and *x2*,*y2*), one *y* value can be interpolated at *x* position through Equation (1)

$$\boldsymbol{y = y1 + (x - x1)(\frac{y2-y1}{x2-x1})} \quad (1)$$

Where:

$$\boldsymbol{slope = (\frac{y2-y1}{x2-x1})} \quad (2)$$

From Equation (1) *n* intermediate synthetic values can be generated, as can be seen in Fig 1 b). Equation (1) is implemented in Python language. Considering 9 synthetic items per pair of original time series values, 61110 synthetic values are generated.

3.3.3 Polynomial interpolation (DA3)

Inspired by linear interpolation, also polynomial interpolation can be used to generate synthetic values. However, unlike linear interpolation which requires at least two points, for polynomial interpolation at least three points are required.

For this type of interpolation, it is necessary to determine the coefficients of the polynomial function, for this, there are different techniques, including the Lagrange method.

Given the *n* points (x_0,y_0), … , (x_{n-1},y_{n-1}), the Lagrange Polynomial is estimated through Eq. (3).

$$\boldsymbol{p(x) = \sum_{i=0}^{n-1} y_i \frac{\prod_{j\neq i}(x-x_j)}{\prod_{j\neq i}(x_i-x_j)}} \quad (3)$$

From Equation (3) the polynomial coefficients are obtained and the polynomial function can be implemented, from it, any point can be estimated, in this case, the synthetics values. The polynomial function is similar to what is shown in Equation (4).

$$\boldsymbol{p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1}} \quad (4)$$

In this work, polynomial interpolation for data augmentation is implemented using the numpy library in Python language through the polyfit function.

Similar to linear interpolation, considering 9 synthetic items per pair of original values, 61110 synthetic values are generated.

Fig 2. shows the results of data augmentation techniques for the first 6 items of the selected pm2.5 time series.

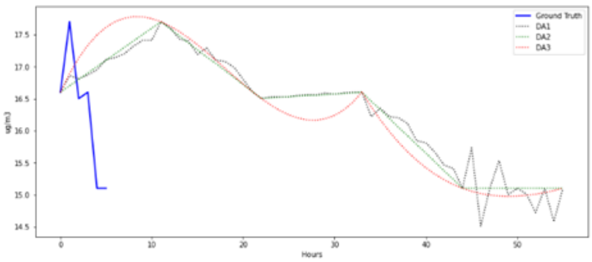

Fig. 3. Augmented pm2.5 time series

3.4 Normalization

Time series scaling has been performed using min-max normalization. Eq. (5) allows to implement min-max scaling between 0 and 1 values.

$$\boldsymbol{x' = \frac{x-\min(x)}{\max(x)-\min(x)}} \quad (5)$$

Where:

$\boldsymbol{x'}$	:	normalized value
$\boldsymbol{x}$	:	value to be normalized
$\boldsymbol{\min(x)}$	:	min value in time series
$\boldsymbol{\max(x)}$	:	max value in time series

3.4 Modeling

In terms of the number of layers, neurons, and dropout rates, the same architecture was used for all the models, they only differed by the types of layers, Table 1 shows the hyperparameters of all architectures. For the compilation of the models, mse was used as the loss function and adam as optimizer, the learning rate of 0.001 was used. For model fitting, 100 epochs and batch_size=500.

Table 1. Hyperparameters of deep learning models

Model	Hyperparameters
LSTM	layers [30, 30, 30, 1]
BiLSTM	activation [relu, relu, relu, sigmoid]
GRU	dropout [0.0, 0.1, 0.1]
BiGRU	

3.5 Evaluation

The results evaluation of the implemented models is carried out in terms of Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE) and R Squared (R^2) that are estimated according to Equations (6), (7) and (8).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pi-Oi)^2}{n}} \quad 6)$$

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

$$R^2 = \sum_{i=1}^{n}(P_i-\bar{O})^2 / \sum_{i=1}^{n}(O_i-\bar{O})^2 \quad (8)$$

RMSE allows to evaluate the results in terms of ug/m^3, while MAPE allows to do it in percentage terms (0-100%), in both cases the best results are those closest to 0. On the other hand, R^2 measures the correlation between the original data and the predicted data, in this case, the best results are those closest to 1.

4. Results and Discussions

In this section, the achieved results are described in detail and discussed with related works.

4.1 Results

Table 3. RMSE of implemented models

Model	None	DA1	DA2	DA3
LSTM	7.0335	**0.7715**	0.3705	0.4583
BiLSTM	**6.7946**	0.7883	0.3651	0.3485
GRU	7.2350	1.2042	**0.1748**	**0.3080**
BiGRU	6.9727	0.9844	0.4430	0.3836
Avg	7.0089	0.9371	**0.3383**	0.3746

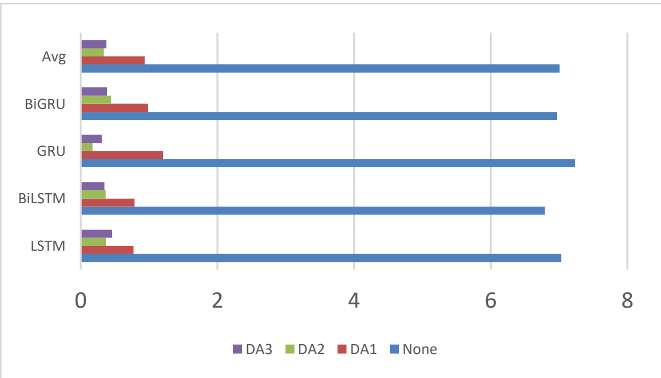

Fig. 4: RMSE

According to Table 3 and Fig. 4, it can be seen that in terms of RMSE, the use of data augmentation techniques allows for greatly improve the error of the implemented models, thus, on average, the models improve between 6.2940 ug/m^3 and 6.6726 ug/m^3. Being BiLSTM the model that improves the most, followed by LSTM, GRU, and BiGRU. According to RMSE the best data augmentation technique is based on linear interpolation (DA2) followed by polynomial interpolation (DA3).

Table 4. MAPE of implemented models

Model	None	DA1	DA2	DA3
LSTM	0.5384	0.0441	0.0365	0.0402
BiLSTM	0.4686	**0.0439**	0.0342	0.0372
GRU	0.5110	0.1342	**0.0193**	**0.0244**
BiGRU	**0.3706**	0.0727	0.0598	0.0314
Avg	0.4721	0.0737	0.0375	**0.0333**

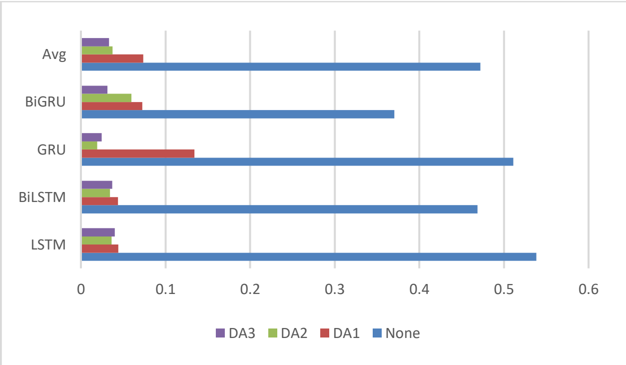

Fig. 5. MAPE

In terms of MAPE, according to Table 4 and Fig. 5, it can be seen that in percentage terms, in the same way as with the RMSE, the models with data augmentation greatly outperform the models without data augmentation. On average, models with data augmentation outperform models without data augmentation between 31.60% and 49.81%. According to MAPE, on average the best data augmentation is DA3 followed by DA2.

Table 5. R^2 of implemented models

Model	None	DA1	DA2	DA3
LSTM	0.5288	**0.9939**	**1.0000**	0.9983
BiLSTM	0.5331	0.9941	0.9988	0.9991
GRU	**0.5368**	0.9922	0.9999	**1.0000**
BiGRU	0.5174	0.9936	0.9996	**1.0000**
Avg	0.5290	0.9936	**0.9996**	0.9993

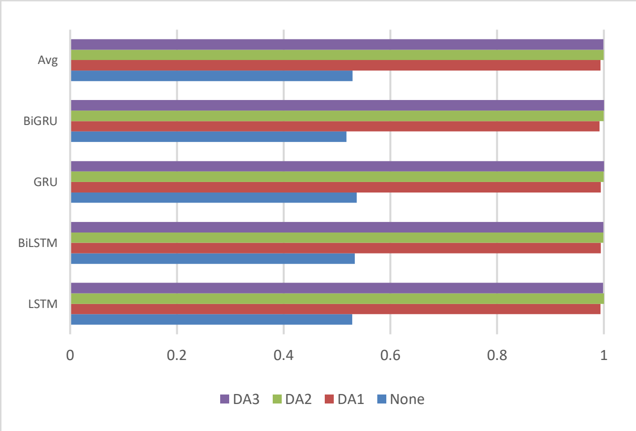

Fig. 6. R^2

Similar to what happened with the RMSE and the MAPE, according to Table 5 and Fig. 6 with the R^2 the superiority of the models with data augmentation is appreciated, surpassing the models without data augmentation on average between 46.12% and 47.98%. According to R^2, the best data augmentation technique is based on linear interpolation (DA2) followed by polynomial interpolation (DA2).

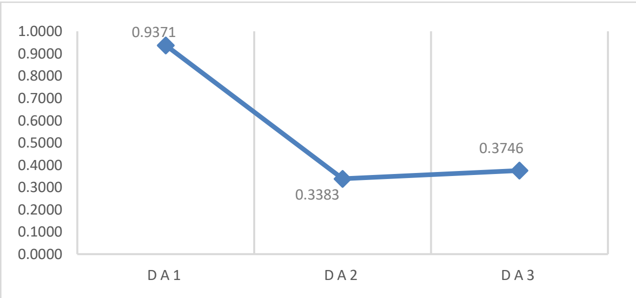

Fig. 7. Average RMSE

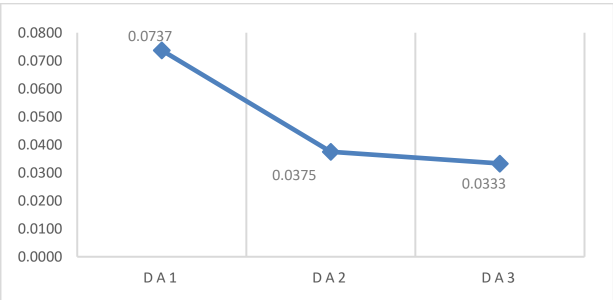

Fig. 8. Average MAPE

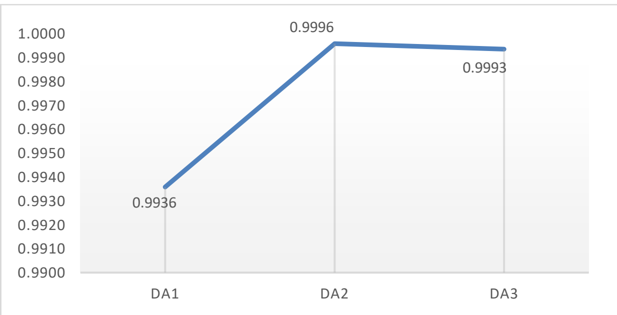

Fig. 9. Average R2

According to the results, it can be seen that there is a significant difference between DA2, and DA3 with respect to DA1.

According to Fig. 7, in terms of RMSE, DA2 beats DA1 by 0.5987ug/m^3, and DA3 by 0.5625ug/m^3. Comparing DA2 with DA3, DA2 outperforms DA3 by 0.0362 ug/m^3. According to Fig. 8, in terms of MAPE, DA2 outperforms DA1 by 3.62%, while DA3 outperforms DA1 by 4.04%. DA2 exceeds DA3 by 0.0362ug/m3. In terms of MAPE, DA3 outperforms DA2 by 0.42%. And, according to Fig. 9, in terms of R^2, DA2 outperforms DA1 by 0.599%, while DA3 outperforms it by 0.576%. comparing DA2 with DA3, the difference is minimal, DA2 exceeds DA3 by 0.024%.

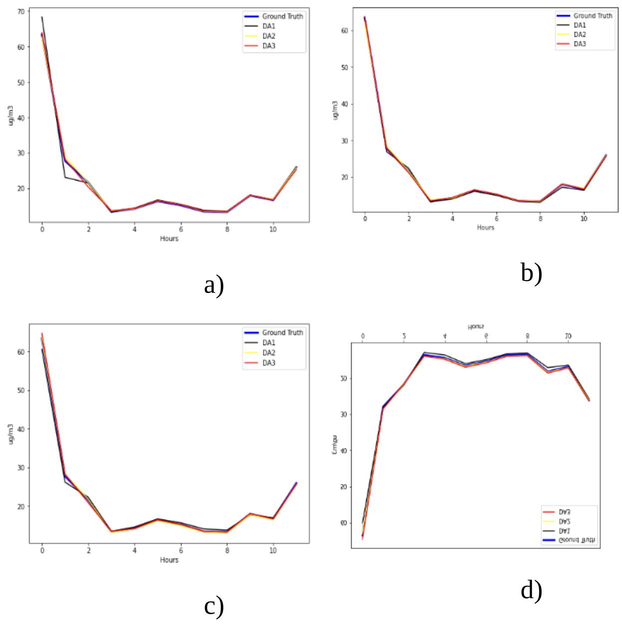

Fig. 10. Comparison of predicted values a) LSTM, b) BiLSTM, c) GRU, and d) BiGRU

Fig. 9 shows graphically predicted values by each implemented model with data augmentation for the first 24 hours of the test data. As can be seen for all implemented deep learning models it is a bit difficult to visually find the difference between each data augmentation technique.

4.2 Discussions

The main difference between DA2 and DA3 is the computational cost due to the higher complexity of the polynomial interpolation compared to the linear interpolation. Considering the training data with 6112 items, the required time to estimate 10, 50, and 100 synthetics values is shown in Table 6 for each pair of pm2.5 time series values. Spyder 4.1.4 is used as IDE with Python 3.8 and AMD Ryzen 7.4700U processor.

Table 6. Required time to generate synthetics items

Interpolation	Synthetic items		
	10	50	100
Linear	0.0218	0.0676	0.1242
Polynomial	1.4145	5.6243	9.2318

Table 7 shows the results achieved and reported by related works. In terms of RMSE, the implemented deep learning models exceed most of the results of the related works. The proposal DA+BiLSTM is only surpassed by the work [24] by Guo et al that uses daily pm2.5 time series with an RMSE=0.2291 ug/m^3, and the work [20] by Flores et al, 2021 that uses linear interpolation for data augmentation with RMSE=0.0556ug/m^3.

Table 7. Comparison with related-works

Work	Technique	Freq.	Train	Test	Metric	Value
Li et al, 2018 [22]	QPSO+SVR	Hourly	1490	638	RMSE	14.05
Liu et al, 2019 [13]	SSHL+LSTM	Daily	1000	100	MAE	
Pak et al, 2020 [14]	LSTM	Daily	10522	2630	RMSE	8.11
Chen et al, 2020 [25]	CHRNN	Hourly			RMSE	4.373
Chu et al, 2020 [23]	CVMD-STASA-SVR	Hourly	360	360	RMSE	3.9409
Zhang et al, 2020 [15]	SSA-LSTM	Hourly	6150	2634	RMSE	6.365
Li et al, 2021[26]	MMAUKF	Hourly		24	RMSE	5.9130
Wang et al, 2021[16]	CNN+LSTM	Hourly	31550	10516	RMSE	14.21
Guo et al, 2021[27]	VMD+IWOA	Daily	572	191	RMSE	0.2291
Nguyen et al, 2021 [19]	ED+LSTM	Hourly	24096	8760	MAE	3.592
Shi et al, 2021 [28]	DSTP-FC	Hourly	6570	2190	RMSE	32.51
Xing et al, 2021 [29]	TDBN	Daily	120	80	RMSE	11.1978
Zhang et al, 2021 [12]	BiLSTM	Hourly	10176	1848	RMSE	6.86
Li et al, 2021[24]	Random Forest	Daily	90%	10%	R^2	0.71
Flores et al, 2021 [20]	DA+LSTM	Hourly	2000	775	RMSE	0.0556
Zheng et al, 2022 [21]	RL+GRU	Hourly	900	300	RMSE	1.8192
Tian et al, 2022 [34]	NN+MOO	Hourly	1500	100	RMSE	11.32
Yin et al, 2022 [30]	DA +ResNet	Hourly			RMSE	18.81
Zhu et al, 2023 [17]	CNN+BiLSTM	Hourly		1150	RMSE	3.88
Proposal	DA1+BiLSTM	Hourly	6112	1529	RMSE	0.7883
	DA2+BiLSTM				RMSE	0.3651
	DA3+BiLSTM				RMSE	0.3485

5. Conclusion

The implemented data augmentation techniques allow to improve all the implemented deep learning models for pm2.5 time series prediction; there is a superiority of DA2 and DA3 over DA1. However, between DA3 and DA2 there is no significant difference in terms of their results, for some models, DA2 produces better results than DA3, and vice versa for other models. The main difference between DA2 and DA3 would lie in the computational cost, DA3 has a higher cost than DA2 since the polynomial interpolation is more complex than the linear interpolation.

Numerals.

References

[1] WHO, "Air pollution." https://www.who.int/health-topics/air-pollution#tab=tab_1 (accessed Jun. 10, 2023).

[2] C. C. Cho, W. Y. Hsieh, C. H. Tsai, C. Y. Chen, H. F. Chang, and C. S. Lin, "In vitro and in vivo experimental studies of PM2.5 on disease progression," *International Journal of Environmental Research and Public Health*, vol. 15, no. 7. 2018, doi: 10.3390/ijerph15071380.

[3] Y. F. Xing, Y. H. Xu, M. H. Shi, and Y. X. Lian, "The impact of PM2.5 on the human respiratory system," *Journal of Thoracic Disease*, vol. 8, no. 1. 2016, doi: 10.3978/j.issn.2072-1439.2016.01.19.

[4] T. Li *et al.*, "PM2.5 exposure associated with microbiota gut-brain axis: Multi-omics mechanistic implications from the BAPE study," *Innovation*, vol. 3, no. 2, 2022, doi: 10.1016/j.xinn.2022.100213.

[5] Z. Xue *et al.*, "ITRAQ based proteomic analysis of PM2.5 induced lung damage," *RSC Adv.*, vol. 9, no. 21, 2019, doi: 10.1039/c9ra00252a.

[6] M. A. H. Suryadhi *et al.*, "Exposure to particulate matter (PM2.5) and prevalence of diabetes mellitus in Indonesia," *Environ. Int.*, vol. 140, 2020, doi: 10.1016/j.envint.2020.105603.

[7] Y. Hao, Y. Zhou, J. Gao, and J. Wang, "A Novel Air Pollutant Concentration Prediction System Based on Decomposition-Ensemble Mode and Multi-Objective Optimization for Environmental System

Management," *Systems*, vol. 10, no. 5, 2022, doi: 10.3390/systems10050139.

[8] I. Nusrat and S. B. Jang, "A comparison of regularization techniques in deep neural networks," *Symmetry (Basel).*, vol. 10, no. 11, 2018, doi: 10.3390/sym10110648.

[9] J. Sijia *et al.*, "Spatial prediction using random forest spatial interpolation with sample augmentation: a case study for precipitation mapping," *Earth Sci. Informatics*, vol. 16, no. 1, 2023, doi: 10.1007/s12145-023-00936-6.

[10] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text Data Augmentation for Deep Learning," *J. Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00492-0.

[11] R. B. Arantes, G. Vogiatzis, and D. R. Faria, "Learning an augmentation strategy for sparse datasets," *Image Vis. Comput.*, vol. 117, 2022, doi: 10.1016/j.imavis.2021.104338.

[12] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, and J. Liu, "Air quality predictions with a semi-supervised bidirectional LSTM neural network," *Atmos. Pollut. Res.*, vol. 12, no. 1, 2021, doi: 10.1016/j.apr.2020.09.003.

[13] Q. Liu, Y. Zou, and X. Liu, "A self-organizing memory neural network for aerosol concentration prediction," *C. - Comput. Model. Eng. Sci.*, vol. 119, no. 3, 2019, doi: 10.32604/cmes.2019.06272.

[14] U. Pak *et al.*, "Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China," *Sci. Total Environ.*, vol. 699, 2020, doi: 10.1016/j.scitotenv.2019.07.367.

[15] Y. Zhang and W. Li, "SSA-LSTM neural network for hourly PM2.5 concentration prediction in Shenyang, China," in *Journal of Physics: Conference Series*, 2021, vol. 1780, no. 1, doi: 10.1088/1742-6596/1780/1/012015.

[16] W. Wang, W. Mao, X. Tong, and G. Xu, "A novel recursive model based on a convolutional long short-term memory neural network for air pollution prediction," *Remote Sens.*, vol. 13, no. 7, 2021, doi: 10.3390/rs13071284.

[17] M. Zhu and J. Xie, "Investigation of nearby monitoring station for hourly PM2.5 forecasting using parallel multi-input 1D-CNN-biLSTM," *Expert Syst. Appl.*, vol. 211, 2023, doi: 10.1016/j.eswa.2022.118707.

[18] L. Zhang, L. Xu, M. Jiang, and P. He, "A novel hybrid ensemble model for hourly PM2.5 concentration forecasting," *Int. J. Environ. Sci. Technol.*, vol. 20, no. 1, 2023, doi: 10.1007/s13762-022-03940-3.

[19] M. H. Nguyen, P. Le Nguyen, K. Nguyen, V. A. Le, T. H. Nguyen, and Y. Ji, "PM2.5 Prediction Using Genetic Algorithm-Based Feature Selection and Encoder-Decoder Model," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3072280.

[20] A. Flores, J. Valeriano-Zapana, V. Yana-Mamani, and H. Tito-chura, "PM2.5 prediction with Recurrent Neural Networks and Data Augmentation," 2021, doi: 10.1109/LA-CCI48322.2021.9769784.

[21] G. Zheng, H. Liu, C. Yu, Y. Li, and Z. Cao, "A new PM2.5 forecasting model based on data preprocessing, reinforcement learning and gated recurrent unit network," *Atmos. Pollut. Res.*, vol. 13, no. 7, 2022, doi: 10.1016/j.apr.2022.101475.

[22] X. Li, A. Luo, J. Li, and Y. Li, "Air Pollutant Concentration Forecast Based on Support Vector Regression and Quantum-Behaved Particle Swarm Optimization," *Environ. Model. Assess.*, vol. 24, no. 2, 2019, doi: 10.1007/s10666-018-9633-3.

[23] J. Chu, Y. Dong, X. Han, J. Xie, X. Xu, and G. Xie, "Short-term prediction of urban PM2.5 based on a hybrid modified variational mode decomposition and support vector regression model," *Environmental Science and Pollution Research*, vol. 28, no. 1. 2021, doi: 10.1007/s11356-020-11065-8.

[24] J. Li *et al.*, "Estimation of ambient PM2.5 in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing," *Environ. Int.*, vol. 151, 2021, doi: 10.1016/j.envint.2021.106445.

[25] Y. C. Chen, T. C. Lei, S. Yao, and H. P. Wang, "Pm2.5 prediction model based on combinational hammerstein recurrent neural networks," *Mathematics*, vol. 8, no. 12, 2020, doi: 10.3390/math8122178.

[26] J. Li, X. Li, K. Wang, and G. Cui, "Atmospheric pm2.5 prediction based on multiple model adaptive unscented kalman filter," *Atmosphere (Basel).*, vol. 12, no. 5, 2021, doi: 10.3390/atmos12050607.

[27] H. Guo, Y. Guo, W. Zhang, X. He, and Z. Qu, "Research on a novel hybrid decomposition–ensemble learning paradigm based on VMD and IWOA for PM2.5 forecasting," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, 2021, doi: 10.3390/ijerph18031024.

[28] P. Shi, X. Fang, J. Ni, and J. Zhu, "An improved attention-based integrated deep neural network for pm2.5 concentration prediction," *Appl. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/app11094001.

[29] H. Xing, G. Wang, C. Liu, and M. Suo, "PM2.5 concentration modeling and prediction by using temperature-based deep belief network," *Neural Networks*, vol. 133, 2021, doi: 10.1016/j.neunet.2020.10.013.

[30] S. Yin, T. Li, X. Cheng, and J. Wu, "Remote sensing estimation of surface PM2.5 concentrations using a deep learning model improved by data augmentation and a particle size constraint," *Atmos. Environ.*, vol. 287, 2022, doi: 10.1016/j.atmosenv.2022.119282.

[31] S. Moritz and T. Bartz-Beielstein, "imputeTS: Time series missing value imputation in R," *R J.*, vol. 9, no. 1, 2017, doi: 10.32614/rj-2017-009.

[32] S. Moritz, “imputeTS,” 2021.

[33] A. Flores, H. Tito-Chura, and H. Apaza-Alanoca, “Data Augmentation for Short-Term Time Series Prediction with Deep Learning,” in *Lecture Notes in Networks and Systems*, 2021, vol. 284, doi: 10.1007/978-3-030-80126-7_36.

[34] Z. Tian and M. Gai, “New PM2.5 forecasting system based on combined neural network and an improved multi-objective optimization algorithm: Taking the economic belt surrounding the Bohai Sea as an example,” *J. Clean. Prod.*, vol. 375, 2022, doi: 10.1016/j.jclepro.2022.134048.

Funding

This research received no external funding and the APC was funded by Universidad Nacional de Moquegua.

Author contributions

Anibal Flores: Conceptualization, Methodology, Software, Field study, Writing-Original draft preparation. **Hugo Tito-Chura:** Data curation, Software, Validation., Field study. **Deymor Centty-Villafuerte:** Field Study, Visualization, Investigation, Writing-Reviewing **Alejandro Ecos-Espino:** Validation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.