AI+ENVIRONMENT SUMMIT

AI-Driven Prediction of Fine Particulate Matter ($PM_{2.5}$) Concentrations Using Attention-Based BiLSTM

Tuyet Nam Thi Nguyen, Tan Dat Trinh.
Saigon University, Ho Chi Minh City, Vietnam.
ntnam@sgu.edu.vn; *trinhtandat@sgu.edu.vn*

1. Introduction

- Fine particulate matter ($PM_{2.5}$), particles with a diameter of less than 2.5 µm, poses a major health risk to humans.
- LSTM & BiLSTM models have been increasingly used to predict $PM_{2.5}$ pollution.
- Attention mechanisms are a powerful enhancement in time-series modeling.
- The number of timesteps directly influences the model's ability to capture temporal dependencies.

2. Research Objectives

- To provide insights into $PM_{2.5}$ pollution in Ho Chi Minh City (HCMC).
- To develop an advanced DL model for enhancing the accuracy of $PM_{2.5}$ concentration prediction.
- To evaluate the influence of different time lag values (i.e., historical input lengths) on model performance.

3. Method

3.1. Data collection

- Hourly concentrations of $PM_{2.5}$ in HCMC: obtained from the AirNow monitoring network.
- Meteorological data: obtained from ERA5.
- Study period: 01/01/2018 to 17/06/2023.

3.2. Data preprocessing

- Data were converted into supervised learning format.

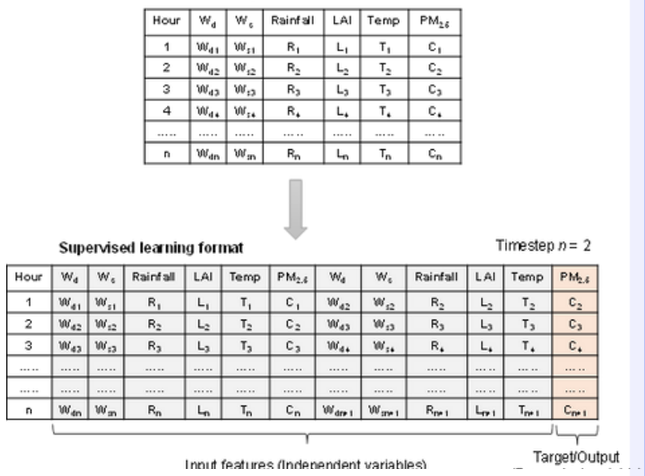

Fig. 1 Data conversion with a timestep of 2 as an example.

3.3. Proposed model

Table 1. Hyperparameters of BiLSTM+Attention model.

Hyperparameters	Value
Input timesteps *n*	2, 6, 12, 18, 24, 36
Number of features of the input dimension	7
Activation function of Conv layer	ReLU
BiLSTM layers	2
Number of features per timestep	256 and 128
Dropout rate	0.5
Number of units of dense layers	128, 64, 1
Optimizer	Adam
Batch size	64
Learning rate	0.001
Number of epochs	100

4. Result & Discussion

4.1. Temporal variation of $PM_{2.5}$ concentrations

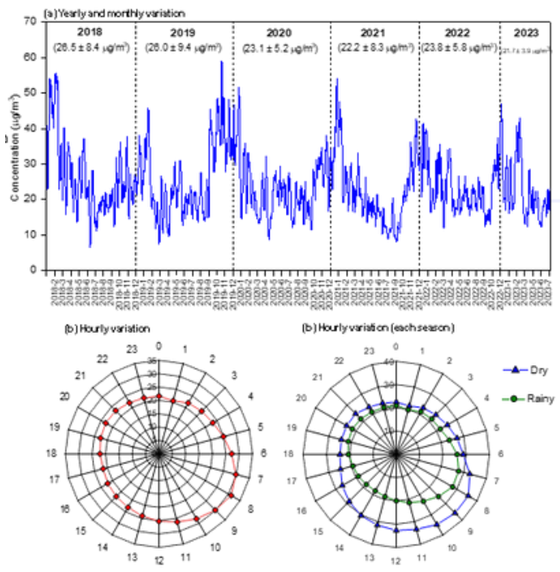

Fig. 2 Concentrations of $PM_{2.5}$

4.2. Effect of meteorological conditions on $PM_{2.5}$ concentrations

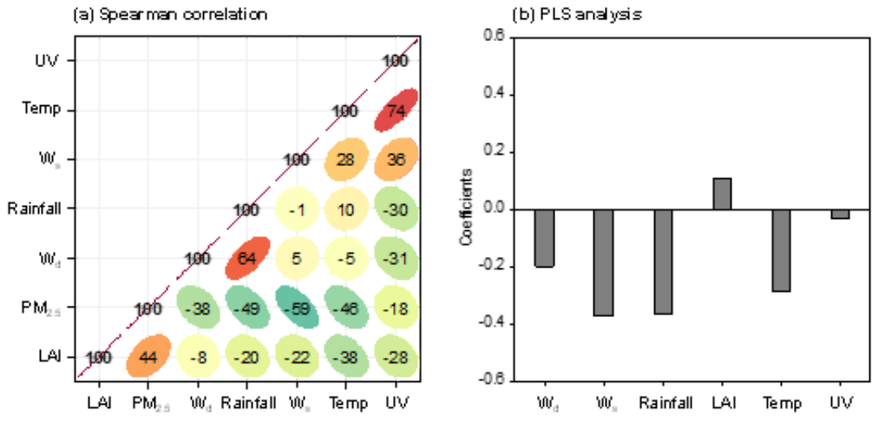

Fig. 3 Spearman and Partial least square (PLS) analysis

4.3. Model performance

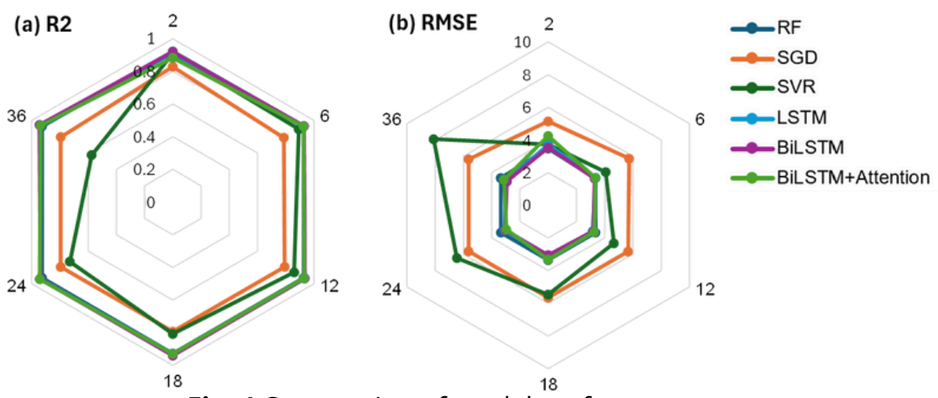

Fig. 4 Comprasion of model performance

Best performance model: BiLSTM with attention, timestep = 24

4.4.Concentration prediction

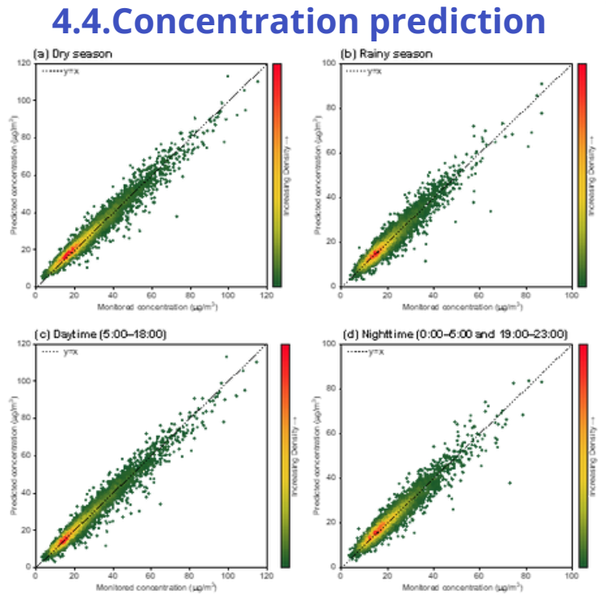

Fig. 5 Scatter plots of monitored and predicted concentrations (BiLSTM+Attention, timestep n=24)

5. Conclusion

- $PM_{2.5}$ concentrations in the rainy season were lower than those in the dry season.
- $PM_{2.5}$ concentrations exhibited strong negative relations with certain meteorological variables (i.e., rainfall intensity, ambient air temperature, and wind speed)
- BiLSTM+Attention achieved its best performance when both auxiliary variables and $PM_{2.5}$ concentrations from the previous 24 hours were used as inputs.
- The model attained a coefficient of determination (R^2) of 0.944 and a root mean square error (RMSE) of 2.957 µg/m³.