

Predicting PM2.5 Concentrations Using Stacking-based Ensemble Model

Haoyuan Zhang^{1,*}, Yilun Jin², Jiaxuan Shi², Shuai Zhang²

¹Guanghua Cambridge International School, Shanghai, China

²School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China

Email address:

zhanghaoyuanc@163.com (Haoyuan Zhang), jinyilun@zufe.edu.cn (Yilun Jin), shijiaxuan@zufe.edu.cn (Jiaxuan Shi),

zhangshuai@zufe.edu.cn (Shuai Zhang)

*Corresponding author

To cite this article:

Haoyuan Zhang, Yilun Jin, Jiaxuan Shi, Shuai Zhang. Predicting PM2.5 Concentrations Using Stacking-based Ensemble Model. *Applied and Computational Mathematics*. Vol. 10, No. 6, 2021, pp. 156-162. doi: 10.11648/j.acm.20211006.14

Received: July 31, 2021; **Accepted:** November 22, 2021; **Published:** December 2, 2021

Abstract: With the increasingly serious air pollution problem, PM2.5 concentration, as an effective indicator to evaluate air quality, has attracted extensive attention from all sectors of society. Accurate prediction of PM2.5 concentrations is of great significance in providing the public with early air pollution warning information to protect public health. With a decade of development, artificial intelligence technology has given birth to various prediction models with high-performance, in particular, brought new impetus to the prediction of PM2.5 concentrations. In this study, a stacking-based ensemble model with self-adaptive hyper-parameter optimization is proposed to solve the PM2.5 concentrations prediction problem. First, the raw data are preprocessed with the normalization method to reduce the influence of the different orders of magnitude of input variables on model performance. Second, the Bayesian optimization method is used to optimize the hyper-parameters of the base predictors to improve their performance. Finally, a stacking ensemble method is applied to integrate the optimized base predictors into an ensemble model for final prediction. In the experiments, two datasets from the air quality stations in different areas are tested with four metrics to evaluate the performance of the proposed model in PM2.5 concentration prediction. The experimental results show that the proposed model outperforms other baseline models in solving the PM2.5 concentrations prediction problem.

Keywords: Ensemble Model, Stacking, Bayesian Optimization, PM2.5 Concentrations, Prediction, Air Quality

1. Introduction

In recent years, along with the rapid development of modern industries, air pollution has become increasingly serious. As one of the major airborne pollutants, PM2.5 has captured considerable focus. PM2.5 refers to the inhalable particulate matter with an aerodynamic diameter of 2.5 μm or less, which is composed of highly active toxic and hazardous substance [1]. If one exposes to high concentrations of PM2.5 for a long time, the normal function of the human body will be affected and the risk of lung cancer will increase [2]. Therefore, to protect public health, providing accurate prediction of PM2.5 concentrations as the early warning information is a matter of urgency.

However, due to the complexity of the PM2.5 formation

and development, how to accurately predict the PM2.5 concentration has become a challenging task. In the past, traditional statistical methods have often been used to predict the PM2.5 concentration [3-5]. Unfortunately, most of these methods assume that there is a linear relationship between historical PM2.5 concentration data and future PM2.5 concentration, with the complex nonlinear relationship between them oversimplified [6]. Recently, benefiting from the booming development of artificial intelligence technology, various high-performance prediction models suitable for describing complex nonlinear relationships have been used to monitor and predict PM2.5 concentrations accurately. In this study, a stacking-based ensemble model with self-adaptive hyper-parameter optimization (SEM-SAHPO) is proposed to predict PM2.5 concentrations accurately. First, after the normalization of data, the proposed model deploys the

Bayesian optimization method to self-adaptively optimize the hyper-parameters of base predictors, including linear regression (LR) [7], support vector regression (SVR) [8], k-nearest-neighbor (KNN) [9], long short-term memory (LSTM) [10], convolutional neural network (CNN) [11], and multi-layer perception neural network (MLP) [12]. Then, the stacking ensemble method is employed to integrate these base predictors. In the experiments, the performance of the proposed model and six baseline models are compared on two datasets from the air quality stations in different areas. The experimental results proved that the proposed model performs better than other baseline models in predicting PM2.5 concentrations.

The remainder of this study is organized as follows. Section 2 reviews the previous works on PM2.5 concentrations prediction and ensemble methods. In Section 3, the methodology of the proposed ensemble model is introduced. Section 4 details the experiments and analyzes the experimental results. Section 5 describes the conclusions and provides the directions of future studies.

2. Related Work

2.1. PM2.5 Concentrations Prediction

The adverse effects of PM2.5 push researchers to predict its concentrations. Earlier studies on the PM2.5 concentrations prediction mainly adopted the statistical methods such as land use regression model [4], auto-regression integrated moving average [5], and Kalman filtering method [13]. The performance of these models based on statistical methods is weak in predicting the extreme points, and has difficulty in dealing with the complex factors involved in the prediction [1]. They usually predict the future PM2.5 concentration by describing the linear relationship between historical PM2.5 concentration data and future PM2.5 concentration, while oversimplifying the complex nonlinear relationship between them [6].

With the persistent development of artificial intelligence technology, the machine learning methods have gradually become popular for air quality prediction, because they can describe the complex nonlinear relationships and greatly improve the prediction accuracy. For example, Gennaro et al. [14] proposed an artificial neural network to predict the PM10 daily concentrations and observed the effectiveness of the artificial neural network in obtaining air quality information. Sinnott and Guan [15] compared the performances of the machine learning methods including LSTM networks and artificial neural networks with the traditional statistical methods for predicting the PM2.5 concentrations of Melbourne. They observed that the machine learning methods are better than the traditional statistical methods, and the LSTM network performs best in solving the PM2.5 concentrations prediction problem. Joharestani et al. [16] adopted several machine learning-based models, including extreme gradient boosting, deep learning, and random forest on multi-source remote

sensing data to predict the PM2.5 concentrations in Tehran's urban area.

Previous literatures have demonstrated that the machine learning methods can predict the PM2.5 concentrations well. In this paper, six machine learning-based prediction models with good prediction performance, including three traditional machine learning models and three deep learning models, are selected as the base predictors to predict the PM2.5 concentrations. To further enhance the accuracy of prediction, the stacking ensemble method is applied to integrate the six base predictors to construct an ensemble model. Moreover, the linear support vector regression (LSVR) [17] is employed as the meta-predictor owing to its excellent flexibility and high efficiency.

2.2. Ensemble Method

In machine learning, the ensemble methods integrating multiple machine learning models have gained considerable attention because they usually outperform single machine learning models. The widely employed ensemble learning methods include bagging, boosting [18], and stacking [19]. Among them, bagging and boosting can be used to construct homogeneous models but not the heterogeneous model; while stacking can be used to construct both homogeneous and heterogeneous models. In addition, stacking is characterized of the flexible structure and stability. These advantages make stacking extensively to be applied for solving various prediction problems, such as bankruptcy prediction [20], user geolocation prediction [21], and earthquake casualty prediction [22].

In recent years, numerous works are emerging in predicting air quality with ensemble methods. For example, Wang and Song [23] developed a deep spatial-temporal ensemble model to predict the air quality in the future and demonstrated that the ensemble model outperforms the single base models and can obtain the more accurate prediction results. Di et al. [24] proposed an ensemble model by integrating multiple machine learning models and predictor variables to predict the PM2.5 concentrations across the contiguous United States. Maciag et al. [25] proposed a clustering-based ensemble model to analyze the PM10 concentrations in London, by evolving spiking neural networks trained on a separate set of time series.

Considering the superior performance of the stacking ensemble method and its successful application in air quality prediction, this study employs the stacking ensemble method to integrate the selected six base predictors. The Bayesian optimization method is employed to optimize the hyper-parameters of the base predictors so that the performance of the proposed model is further improved.

3. Methodology

3.1. Dataset Exploration and Preprocessing

The datasets used in this study involve the PM2.5 concentrations records detected by two air quality stations in

different areas. In detail, the first one [26] is the hourly PM_{2.5} concentrations data from Beijing Olympic Sports Center in China, which is named as BOSC in this study. It includes 4416 records detected between 00:00 h on 1 March and 23:00 h on 1 August 2013. The other one [27] is the hourly PM_{2.5} concentrations data in the Xuhui site of Shanghai, China, which is named as XHSH in this study. It includes 2208 records detected between 00:00 h on 1 October and 22:00 h on 31 December 2015.

To improve the availability and standardization of dataset,

the raw data should be preprocessed. In this study, the raw PM_{2.5} concentrations are preprocessed using the normalization method to reduce the influence of the different orders of magnitude of input variables on model performance.

3.2. The Proposed SEM-SAHPO Model

The sub-section describes the proposed SEM-SAHPO model. The framework of the proposed model is shown in Figure 1 and the details are as follows.

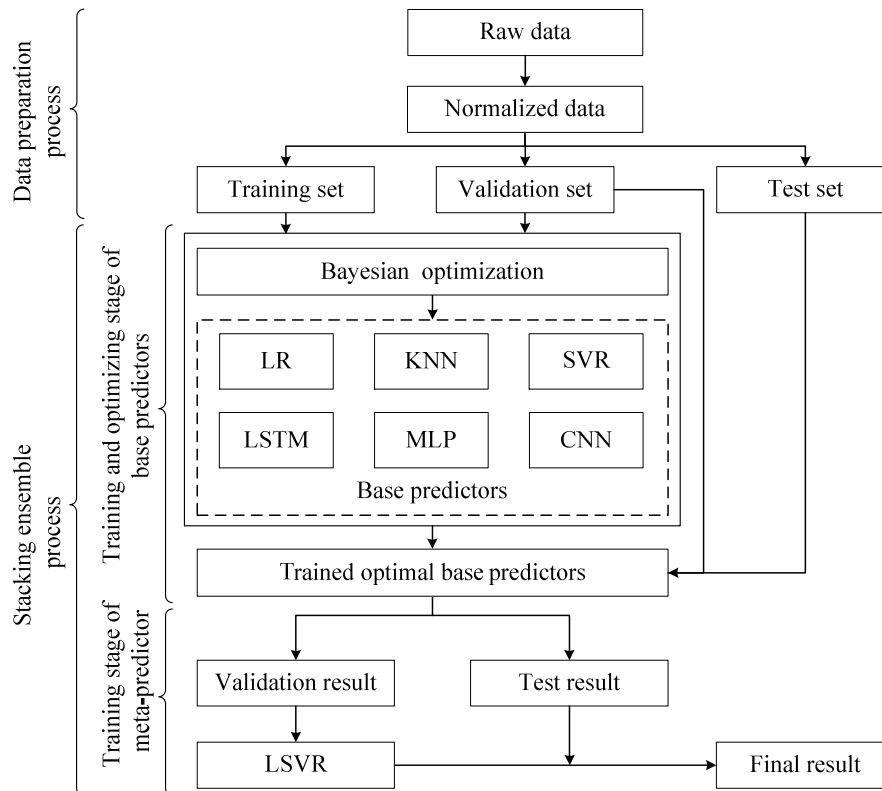


Figure 1. Framework of the proposed model.

3.2.1. Bayesian Optimization Method

In staking-based ensemble models, hyper-parameter setting of the base predictors influences the prediction accuracy of the base predictors and further impacts the ensemble model's performance. This necessitates the optimization of the hyper-parameters of base predictors. Traditional hyper-parameter optimization methods, (e.g., the grid search method and trial-error method), realize the hyper-parameter optimization by searching the whole hyper-parameter space, which is exponentially expensive and inefficient [28]. The emergence of the Bayesian optimization method provides a promising alternative for solving the hyper-parameter optimization problem [29, 30]. Applying the Bayesian optimization method to optimize hyper-parameters can get high-quality solutions with high efficiency [30]. Therefore, the Bayesian optimization method is employed in this study to automate the optimal configuration of the hyper-parameters of base predictors. Through the Bayesian optimization method, the performance of the ensemble model can be enhanced.

3.2.2. Stacking Ensemble Method

The stacking ensemble method is a promising ensemble learning strategy that integrates multiple base predictors into an ensemble model, and characterized of stronger robustness and better performance [31]. In this study, to obtain the higher prediction accuracy, a stacking-based ensemble model is proposed for PM_{2.5} concentrations prediction. Six popular machine learning-based models with good prediction ability, including LR, KNN, SVR, LSTM, MLP, and CNN, are employed as the base predictors. The LSVR with excellent flexibility and high efficiency is selected as the meta-predictor.

The proposed ensemble model includes two processes, i.e., data preparation process and stacking ensemble process. The stacking ensemble process consists of two stages, i.e., the training and optimizing stage of base predictors and the training stage of meta-predictor. In the data preparation process, the raw data are normalized, and the obtained normalized data are divided into three parts, including

training set, validation set, and test set. In the training and optimizing stage of base predictors, the Bayesian optimization method is firstly applied to optimize the hyper-parameters of six base predictors based on the training set and validation set. Then the six base predictors are trained on the training set to obtain the trained optimal base predictors. In the training stage of meta-predictor, the six trained optimal base predictors are firstly used to predict the validation set. Then the six validation results obtained by the trained optimal base predictors are concatenated into one feature matrix, which are employed as the input to train the meta-predictor.

For testing the trained meta-predictor and obtaining the final prediction results, the six trained optimal base predictors are tested on the test set first. Then the six test results of trained optimal base predictors are concatenated into one feature matrix, which are employed as the input to test the meta-predictor to obtain the final results.

4. Experiment

This section introduces the statistical metrics for evaluating all models' performance and analyzes and discusses the experimental results of all models in predicting the PM2.5 concentrations. All models and methods were implemented with Python programming language. The experimental computer contained an Intel Core i7-8700 3.2 GHz CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 2080 GPU.

4.1. Evaluation Metrics

Four widely used statistical metrics are adopted to evaluate the models' performance, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination

(R^2). The calculation process of these four metrics is defined as follows [32, 33]:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{x}_i - x_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (2)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{x}_i - x_i}{x_i} \right| \quad (3)$$

$$R^2 = \left[1 - \frac{\sum_{i=1}^N (\hat{x}_i - \bar{x})^2}{\sum_{i=1}^N (\bar{x} - x_i)^2} \right] \quad (4)$$

where $x_i (1 \leq i \leq N)$ and $\hat{x}_i (1 \leq i \leq N)$ denote the observed and predicted PM2.5 concentration at time t , respectively, where N denotes the size of test samples. \bar{x}_i denotes the average PM2.5 concentration of N test samples. In particular, for MAE, RMSE, and MAPE, the lower value indicates the better prediction performance. For R^2 , the higher value represents the better prediction performance.

4.2. Experimental Results Analysis

Table 1 shows the evaluation results obtained by the proposed SEM-SAHPPO model and other baseline models on BOSC and XHSH datasets. Moreover, to present the evaluation results more intuitively, the predicting results of all models are depicted in column charts, as shown in Figure 2.

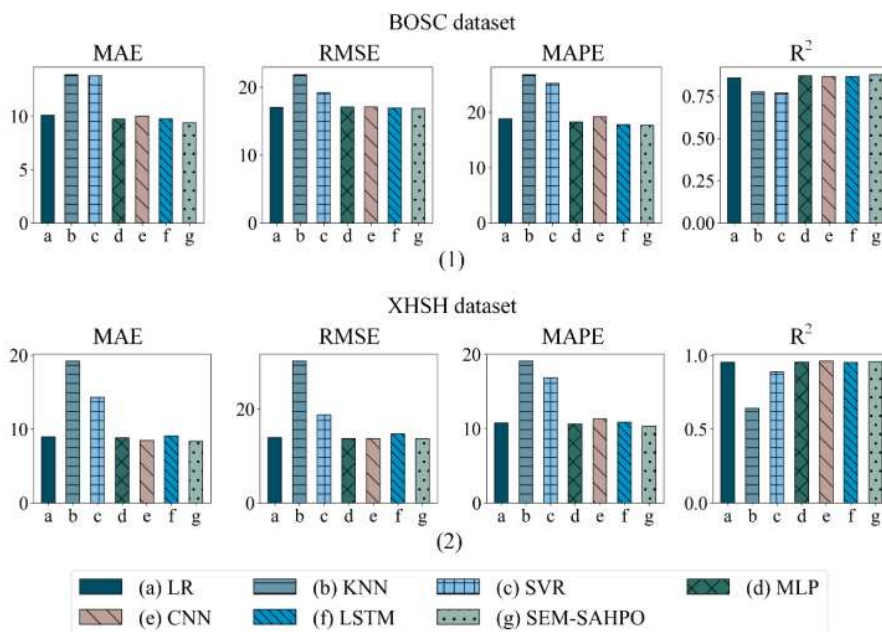


Figure 2. Evaluation results of seven models on BOSC and XHSH datasets.

Table 1. Evaluation results of seven models on BOSC and XHSH dataset.

Dataset	Model	Evaluation metrics			
		MAE	RMSE	MAPE (%)	R ²
BOSC	LR	10.0874	17.0348	18.7967	0.8571
	KNN	13.9017	21.8244	26.7727	0.7758
	SVR	13.7774	19.2143	25.2255	0.7675
	MLP	9.7502	17.0759	18.2392	0.8687
	CNN	10.0316	17.0617	19.1226	0.8667
	LSTM	9.7596	16.9246	17.7884	0.8655
	SEM-SAHPO	9.4088	16.8463	17.6212	0.8765
XHSH	LR	8.9539	13.9309	10.7290	0.9539
	KNN	19.1787	30.1867	19.0619	0.6408
	SVR	14.3092	18.7772	16.8485	0.8892
	MLP	8.8058	13.7244	10.6075	0.9551
	CNN	8.4616	13.6999	11.3042	0.9608
	LSTM	9.0867	14.7388	10.8548	0.9514
	SEM-SAHPO	8.3605	13.7253	10.3332	0.9589

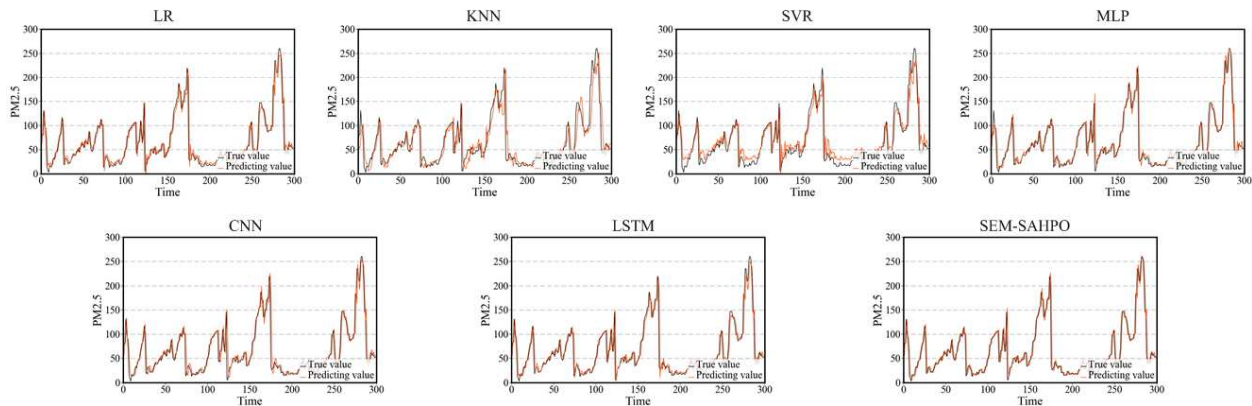
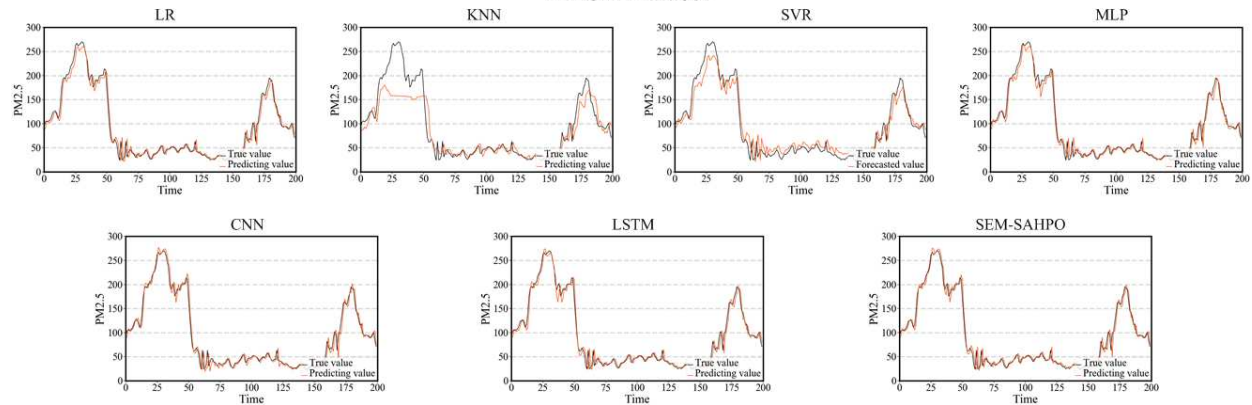
Note: Significant values are boldfaced.

The observed results from Table 1 and Figure 2 can be summarized below:

- (1) The deep learning models (i.e., MLP, CNN, and LSTM) realize better predicting results than those traditional machine learning models (i.e., LR, KNN, and SVR). For example, in terms of the BOSC dataset, the MAE obtained by LR, KNN, and SVR respectively (i.e.,

10.0874, 13.9017, and 13.7774) are higher than the MAE obtained by MLP, CNN, and LSTM respectively (i.e., 9.7502, 10.0316, and 9.7596). In terms of the XHSH dataset, the R² obtained by LR, KNN, and SVR are 0.9539, 0.6408, and 0.8892, respectively, whereas the R² obtained by MLP, CNN, and LSTM are 0.9551, 0.9608, and 0.9514, respectively. This demonstrates that compared to the traditional machine learning models with simple structure, the deep learning models are more suitable for processing PM2.5 concentration data.

- (2) The proposed SEM-SAHPO model outperforms deep learning models (i.e., MLP, CNN, and LSTM) and traditional machine learning models (i.e., LR, KNN, and SVR), because SEM-SAHPO model can integrate the strengths of these base predictors to obtain more accurate and robust results. For example, in terms of the XHSH dataset, the MAPE of MLP, CNN, and LSTM (i.e., 10.6075, 11.3042, and 10.8548) are higher than the MAPE of SEM-SAHPO (i.e., 10.3332). To sum up, the SEM-SAHPO model achieves the better predicting results than other baseline models in most metrics, thus it can be concluded that the SEM-SAHPO model is more effective in capturing the complicated variation patterns of PM2.5 concentration.

BOSC Dataset**Figure 3.** Predicting and true PM2.5 concentration of seven models on BOSC dataset.**XHSH Dataset****Figure 4.** Predicting and true PM2.5 concentration of seven models on XHSH dataset.

To further verify the prediction accuracy of the proposed SEM-SAHPO model, the predicting values of all models are visualized. Figure 3 and Figure 4 show the predicting PM_{2.5} concentration and true PM_{2.5} concentration on BOSC dataset and XHSH dataset, respectively. It can be found that SEM-SAHPO model capture the inherent trend of PM_{2.5} concentration more accurately than other baseline models. Furthermore, SEM-SAHPO model is capable of generating smoother predicting results when the PM_{2.5} concentration fluctuates frequently. Therefore, the conclusion can be drawn that the SEM-SAHPO model outperforms other baseline models in accuracy and stability.

5. Conclusion and Future Work

Accurately predicting the PM_{2.5} concentrations is of great significance for providing early warning information of air pollution and addressing the corresponding public health problems. In this study, an ensemble model is proposed to predict the PM_{2.5} concentrations. Specifically, the Bayesian optimization method is applied to optimize the hyper-parameters of base predictors automatically, and the stacking ensemble method is adopted to integrate the optimized base predictors for achieving the cooperative advantages. The proposed ensemble model was verified on two PM_{2.5} concentrations datasets based on four evaluation metrics. The experimental results demonstrated that the proposed model outperforms other baseline models.

Although the proposed SEM-SAHPO model in this study can obtain promising predicting results, some aspects can be improved in the future. Firstly, the employed LSVR can be extended with some advanced heuristic search algorithms (e.g., genetic algorithm and particle swarm optimization algorithm) to shorten the running time of the proposed model and further improve the model's efficiency. Secondly, the other ensemble methods (e.g., bagging method and boosting method) can be explored to improve the structure of the proposed model to further enhance its performance. Thirdly, more evaluation indicators can also be considered to evaluate the performance of the predictors, so as to obtain more comprehensive evaluation results. In addition, the proposed model can also be applied to predict other air pollutants concentration or to solve other forecasting tasks in other fields, such as traffic flow predicting problem, wind speed predicting problem, and stock price prediction problem.

References

- [1] Sun, W., and Sun, J. Y. (2017). Daily PM_{2.5} concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *Journal of Environmental Management*, 188, 144-152.
- [2] Song, Y. Z., Yang, H. L., Peng, J. H., Song, R. Y., Sun, Q., and Li, Y. (2015). Estimating PM_{2.5} concentrations in Xi'an city using a generalized additive model with multi-source monitoring data. *PLoS One*, 10 (11), e0142149.
- [3] Lee, H. J., Chatfield, R. B., and Strawa, A. W. (2016). Enhancing the applicability of satellite remote sensing for PM_{2.5} estimation using MODIS deep blue AOD and land use regression in California, United States. *Environmental Science & Technology*, 50 (12), 6546-6555.
- [4] Hystad, P., Setton, E., Cervantes, A., Poplawski, K., Deschenes, S., Brauer, M., et al. (2011). Creating national air pollution models for population exposure assessment in Canada. *Environmental Health Perspectives*, 119 (8), 1123-1129.
- [5] Jian, L., Zhao, Y., Zhu, Y. P., Zhang, M. B., and Bertolatti, D. (2012). An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. *Science of the Total Environment*, 426, 336-345.
- [6] Wang, X. P., and Sun, W. B. (2019). Meteorological parameters and gaseous pollutant concentrations as predictors of daily continuous PM_{2.5} concentrations using deep neural network in Beijing-Tianjin-Hebei, China. *Atmospheric Environment*, 211, 128-137.
- [7] Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association*, 83 (402), 414-425.
- [8] Smola, A., and Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155-161.
- [9] Hart, P. (1968). The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, 14, 515-516.
- [10] Chen, J., Zeng, G. Q., Zhou, W. N., Du, W., and Lu, K. D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Conversion and Management*, 165, 681-695.
- [11] Ferreira, A., and Giralaldi, G. (2017). Convolutional neural network approaches to granite tiles classification. *Expert Systems with Applications*, 84, 1-11.
- [12] Liu, D., Wang, J. L., and Wang, H. (2015). Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. *Renewable Energy*, 78, 599-608.
- [13] Djalalova, I., Monache, D. L., and Wilczak, J. M. (2015). PM_{2.5} analog forecast and Kalman filter post-processing for the community multiscale air quality (CMAQ) model. *Atmospheric Environment*, 119, 431-442.
- [14] Gennaro, G. D., Trizio, L., Gilio, A. D., Pey, J., Perez, N., Cusack, M., et al. (2013). Neural network model for the prediction of PM₁₀ daily concentrations in two sites in the Western Mediterranean. *Science of the Total Environment*, 463-464, 875-883.
- [15] Sinnott, R. O., and Guan, Z. (2018). Prediction of air pollution through machine learning approaches on the cloud. In *Proceedings of 2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, Zurich, Switzerland, December 17, pp. 51-60.
- [16] Joharestani, M. Z., Cao, C., Ni, X., Bashir, B., and Talebiesfandarani, S. (2019). PM_{2.5} prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere*, 10 (7), 373.

- [17] Ho, H. C., and Lin, C. J. (2012). Large-scale linear support vector regression. *Journal of Machine Learning Research*, 13, 3323-3348.
- [18] Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5 (2), 197-227.
- [19] Wolpert, D. (1992). Stacked generalization. *Neural Networks*, 5 (2), 241-259.
- [20] Fedorova, E., Gilenko, E., and Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40 (18), 7285-7293.
- [21] Han, B., and Cook, P. (2013). A stacking-based approach to twitter user geolocation prediction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Sofia, Bulgaria, August 4-9, pp. 7-12.
- [22] Cai, S. Z., Yin, Y. Q., Wang, D. J., Li, Z. W., and Wang, Y. Z. (2021). A stacking-based ensemble learning method for earthquake casualty prediction. *Applied Soft Computing*, 101, 107038.
- [23] Wang, S. J., and Song, G. J. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, 314, 198-206.
- [24] Di, Q., Amini, H., Shi, L. H., Kloog, I., Silvern, R., Kelly, J., et al. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment International*, 130, 104909.
- [25] Maciag, P. S., Kasabov, N., Kryszkiewicz, M., and Bembenik, R. (2019). Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area. *Environmental Modelling & Software*, 118, 262-280.
- [26] Zhang, S. Y., Guo, B., Dong, A. I., He, J., Xu, Z. P., and Chen, X. S. (2017). Cautionary tales on air-quality improvement in Beijing. *Proceedings of the Royal Society A: Mathematical, Physical, and Engineering Sciences*, 473 (2205), 20170457.
- [27] Liang, X., Li, S., Zhang, S. Y., Huang, H., and Chen, X. S. (2016). PM2.5 Data reliability, consistency and air quality assessment in five Chinese cities. *Journal of Geophysical Research Atmospheres*, 121 (17), 10220-10236.
- [28] Nguyen, V., Rana, S., Gupta, S., Li, C., and Venkatesh, S. (2017). Budgeted batch bayesian optimization with unknown batch sizes. *arXiv preprint arXiv: 1703.04842*.
- [29] Fröhlich, L. P., Klenske, E. D., Daniel, C. G., and Zeilinger, M. N. (2019). Bayesian optimization for policy search in high-dimensional systems via automatic domain selection. In *Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, November 3-8, pp. 757-764.
- [30] Xia, F. Y., Liu, C. Z., Li, Y. Y., and Liu, N. N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78, 225-241.
- [31] Cai, T. A., H, H. L., and Zhang, W. Y. (2018). Breast cancer diagnosis using imbalanced learning and ensemble method. *Applied and Computational Mathematics*, 7 (3), 146-154.
- [32] Zhang, C., Zhou, J. Z., Li, C. S., Fu, W. L., and Peng, T. (2017). A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Conversion and Management*, 143, 360-376.
- [33] Rodrigues, F., Markou, I., and Pereira, F. C. (2019). Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach. *Information Fusion*, 49, 120-129.