INTERNATIONAL JOURNAL OF CONSERVATION SCIENCE

ISSN: 2067-533X Volume 14, Issue 4, October-December 2023: 1619-1626 www.ijcs.ro

DOI: 10.36868/IJCS.2023.04.24

ENHANCING ECOSYSTEM BIODIVERSITY THROUGH AIR POLLUTION CONCENTRATIONS PREDICTION USING SUPPORT VECTOR REGRESSION APPROACHES

Syaidatul Umairah SOLEHAH[1], Aida Wati Zainan Abidin[1], Saiful Nizam WARRIS[2], Wan Nur SHAZIAYANI[1], Balkish Mohd OSMAN[1], Nurain IBRAHIM[1], Norazian Mohamed NOOR[3], Ahmad Zia UL-SAUFIE[1,*]

[1] School of Mathematical Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA,40450 Shah Alam, Selangor, Malaysia.

[2] School of Computing Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA,13500 Permatang Pauh, Penang, Malaysia.

[3] Faculty of Civil Engineering & Technology, Centre of Excellence Geopolymer and Green Technology (CEGEoGTech), Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis, Malaysia

Abstract

Air is the most crucial element for the survival of life on Earth. The air we breathe has a profound effect on our ecosystem biodiversity. Consequently, it is always prudent to monitor the air quality in our environment. There are few ways can be done in predicting the air pollution index (API) like data mining. Therefore, this study aimed to evaluate three types of support vector regression (linear, SVR, libSVR) in predicting the air pollutant concentration and identify the best model. This study also would like to calculate the API by using the proposed model. The secondary daily data is used in this study from year 2002 to 2020 from the Department of Environment (DoE) Malaysia which located at Petaling Jaya monitoring station. There are six major pollutants that have been focusing in this work like PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, and O_3. The root means square error (RMSE), mean absolute error (MAE) and relative error (RE) were used to evaluate the performance of the regression models. Experimental results showed that the best model is linear SVR with average of RMSE = 5.548, MAE = 3.490, and RE = 27.98% because had the lowest total rank value of RMSE, MAE, and RE for five air pollutants (PM_{10}, $PM_{2.5}$, SO_2, CO, O_3) in this study. Unlikely for NO_2, the best model is support vector regression (SVR) with RMSE = 0.007, MAE = 0.006, and RE = 20.75% in predicting the air pollutant concentration. This work also illustrates that combining data mining with air pollutants prediction is an efficient and convenient way to solve some related environment problems. The best model has the potential to be applied as an early warning system to inform local authorities about the air quality and can reliably predict the daily air pollution events over three consecutive days. Besides, good air quality plays a significant role in supporting biodiversity and maintaning healthy ecosystems.

***Keywords**: Aair pollution index; Data mining; Support vector regression; Linear SVR; LibSVR*

Introduction

Air pollution is widely acknowledged as a serious public health hazard on a global scale. Malaysia, like other developing countries in Southeast Asia, is battling with urban air pollution because of growing industrialization and urbanization, which is especially problematic in

* Corresponding author: ahmadzia101@uitm.edu.my

Malaysia. Increase of air pollutant would negatively impacts on biodiversity including changes in species composition, loss of habitat and food sources, water and soil contamination and genetic diversity reduction. Industrial, energy-generating, and transportation- related exhaust emissions, as well as forest fires and open burning in open spaces, are significant sources of air pollution [1]. Travel, electricity, and fuel consumption are all expanding in lockstep with population growth and have developed into important challenges. As a result, air quality monitoring and evaluation required as it is important to keep an eye on and control air pollution, as well as try to avoid the negative effects of air pollution caused by health problems.

Therefore, many countries develop regulations to reduce air pollution for the long-term protection of the environment [2]. Recent years have seen a rapid increase in research into air quality issues, indicating that people are paying more attention to these issues. Air quality research has primarily concentrated on three areas: air quality monitoring, air quality causal analysis, and air quality forecasting and prediction. Good air quality has several impacts on biodiversity including increase of forest health, reduce respitory stress on wildlife, reduce pollinator disruption and ensure the wellbeing of its ecosystems and people. *D. Zhang and S.S. Woo* [3] expressed that the air quality data can be obtained easily by installing environmental sensors or by downloading it from open databases. Along with the government-built air quality monitoring stations, numerous low-cost, portable, and lightweight air quality monitors have been invented.

As a result, some researchers have done research to ascertain the elements that contribute to air pollution. *H. Baltaci et al.* [4] reported that there are also other studies on air quality prediction that use a variety of methodologies, such as the use of neural networks and support vector regression (SVR) to forecast three days in advance the concentrations of SO_2, CO, and PM_{10} in Istanbul's Besiktas District. *C. Puspitasari et al.* [5] used the support vector regression machine with different kernels to predict the hourly O_3 concentration in the city of Surabaya and compared the prediction result with artificial neural networks, while *A. Suárez Sánchez et al.* [6] established an air quality regression model in the urban area of Aviles, Spain using support vector regression technology based on experimental data of air pollutants from 2006 to 2008. In addition, *W. Yang et al.* [7] predicted hourly $PM_{2.5}$ in the China urban atmospheric environment using support vector regression.

Data mining predictive model especially support vector regression (SVR) can handle both linear and non-linear data. This approach is capable of dealing with multicollinearity and outliers and perform well with big training datasets. It is not only capable of overcoming the limitations of conventional prediction techniques in small sample and high-dimensional application scenarios, but also has superior generalization performance [8]. Therefore, the support vector regression machine has been effectively implemented in the model for predicting air quality. *H. Karimian et al*. [9] stated that data mining methods can be employed to analyze and extract information from existing Air Pollution Index (API) data for predictive purposes. The recent advancement in data mining algorithms has facilitated numerous successful projects across diverse applications, particularly in the realm of air pollution data.

Thus, to fill the gap, the aim of this paper is to compare the three types of support vector regression (linear, SVR, libSVR) in predicting the air pollutant concentration to enhance the biodiversity especially in Malaysia. This study also would like to identify the best model of support vector regression in predicting the air pollutant concentration.

Methodology

This location of this study was in Petaling Jaya because of its rapid population growth and geographic significance, with an area of approximately 97.2 square kilometres (37.5sq.mi). Petaling Jaya is prominently located at one of the Klang Valley's urban centres and is surrounded by the Malaysian capital, Kuala Lumpur.

This study used secondary data that obtained from Department of Environment (DOE) from 2002 to 2020 for Petaling Jaya monitoring station. The observations for this dataset are 6852 observations. However, the availability of data for $PM_{2.5}$ only started from July 2017 to 2020. The variables in this dataset were six major air pollutants (in concentration) like sulphur dioxide (SO_2), particulate matter ($PM_{2.5}$, PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), and carbon monoxide (CO) and others like wind speed, wind direction, temperature, and humidity. In addition, the data of each variable is in daily average. The missing data will be handled by using Expectation Maximization (EM) [10].

Support Vector Regression (SVR) is a support vector algorithm for regression problems [11]. It can be adept at managing both linear and non-linear data flawlessly. Three key concepts associated with SVR include margin, hyperplane, and support vectors. A hyperplane is a plane that separates the given data points into two distinct areas. SVR aims to identify the ideal hyperplane, ensuring that the discrepancies between the training data and the hyperplane do not exceed ε. In other words, it minimizes the gap between the support vectors on either side of the hyperplane [8, 2].

Therefore, a regression function f(x) can be obtained by training SVR so that its deviation from the actual value y is not more than ε for each training vector x. The regression function f(x) can be obtained as:

$$f(x) = \sum_{i=1}^{m} (a_i^* - a_i) K(x, x_i) + b \tag{1}$$

where: x is the input feature vector, b is the bias term, a_i^* and a_i are the Lagrange multipliers and $K(x, x_i)$ is a kernel function. Finally, since most realistic problems are non - linear, the kernel trick is commonly performed by mapping training data onto the high-dimensional feature space. Therefore, the kernel function forms that can be obtained from the following equations:

Linear kernel function,

$$K(x, x_i) = x^T x_i \tag{2}$$

Radial basis kernel function,

$$K(x, x_i) = (\gamma x^T x_i + 1)^2, d \in (1, n), \gamma > 0 \tag{3}$$

This study includes several types of SVR like linear and libSVR since it is regression tasks. Linear Support Vector Machine based upon the JMySVM. This operator uses the Java implementation of Stefan Rüping's support vector regression mySVR [13]. It is limited to the dot (linear) kernel, but generates a high-performance model containing only the linear coefficient for quicker model deployment.

LibSVR is an integrated software for regression (epsilon-SVR, nu-SVR). In nu-SVR, it is utilised to select the fraction of support vectors that wish to retain relative to the total number of samples in your dataset [14]. In epsilon-SVR, the parameter is incorporated into the formulation of the optimization problem and determined automatically (optimally) for the dataset.

Performance indicators in this research work are used to determine accuracy and errors of support vector regression (linear, SVR, libSVR). This was accomplished by performing a computation utilizing a performance indicator to determine the error value and ensuring that the best fit condition could be applied to all variables of air pollution stated. The best fit condition exists when the calculation produces the lowest error percentage. The indicators used to identify the best method for prediction of API are the root mean square error (RMSE), relative error (RE) and, mean absolute error (MAE).

The results of RMSE, MAE, and RE for the prediction model will be compared to determine the best model. The lowest value of RMSE, MAE, and RE will be ranked as one, and the second lowest value will be ranked as two, and the highest is ranked as three. The best support vector regression (linear, SVR, libSVR) model for API prediction will indicates from the lowest total ranking values of RMSE, MAE, and RE as stated by *A.Z.U. Saufie et al.* [12].

Results and discussion

The characteristics of the air pollutants can be determined by using the central tendency, measure of dispersion and data visualization. From the table 1, it reveals that all the variables are not normally distributed because the skewness is not equal to zero. Besides, from the skewness results, it indicates that SO_2, $PM_{2.5}$, PM_{10}, O_3, NO_2, CO, wind direction, and temperature are skewed to the right because the skewness values are positive. However, wind speed and humidity are skewed to the left due to the negative values. Thus, the reason for using the support vector regression (SVR) in this study is because this approach does not require assumptions of normality. In addition, PM_{10} has the highest standard deviation which means that the PM_{10} concentrations are far from the mean of set and spread over. This indicates that PM_{10} has a wobbling dataset compared to other pollutants.

Table 1. Descriptive statistics

Variables	Mean	Median	Standard Deviation	Skewness	Maximum
SO_2 (ppm)	0.0041	0.0038	0.0026	1.2110	0.0235
$PM_{2.5}$ (μg/m^3)	24.8251	22.3441	12.6394	4.2180	140.2732
PM_{10} (μg/m^3)	48.7529	44.2396	24.4902	4.3590	482.2083
O^3 (ppm)	0.0138	0.0129	0.0074	1.5840	0.0869
NO^2 (ppm)	0.0289	0.0284	0.0083	0.2740	0.0657
CO (ppm)	1.3505	1.2755	0.4714	1.2850	6.6875
Wind Speed (mph)	3.5991	3.8583	1.6535	- 0.3620	12.3833
Wind Direction	2.7362	2.6253	0.7939	0.5040	6.2442
Temperature (°C)	28.1907	28.1766	1.8283	0.2380	35.7707
Relative humidity (%)	73.5145	74.2500	7.8515	- 0.0516	95.8750

Prediction Model

This section presents the results of prediction which are the root mean squared error (RMSE), mean absolute error (MAE), and relative error (RE) in predicting the air pollutants concentration. The best proposed model is chosen based on the lowest RMSE, MAE, and RE. To make it easier, the RMSE, MAE, and RE was ranked as one which is the best performance, followed by two and then three. Next, the rank will be summed to choose the best proposed model. Moreover, the next day prediction will be representing as day one, the next 2 days prediction will be presenting as day two, and the next 3 days prediction will be presenting as day three.

Particulate Matter (PM_{10})

PM_{10} concentration is predicted by using three models of support vector machine in regression (SVR, LibSVR, Linear). Therefore, table 2 shows the results of RMSE, MAE, and RE in predicting the PM_{10} concentration by using those methods. Prediction for day 1, it reveals that linear SVR is the best proposed model since it has lowest RMSE, MAE and RE. Along with day 1, prediction of day 2 and day 3 also reveals that the linear SVR is the best proposed model due to lowest value of RMSE, MAE and RE. Nevertheless, as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE.

Carbon Monoxide (CO)

The performance of support vector machine in regression for predicting the CO concentration (SVR, LibSVR, Linear) is shown in table 3. Prediction for day 1,2, and 3 shows that linear SVR is the best proposed model since it has lowest RMSE, MAE and RE. As well as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE. Moreover, SVR and LibSVR is the second and third best model in predicting the CO concentration.

Table 2. Model Evaluation of PM_{10} Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	15.789	9.345	19.72%	2	1	2	5
	LibSVR	20.977	9.895	19.87%	3	3	3	9
	Linear	14.572	9.424	19.28%	1	2	1	4
2	SVR	21.486	11.705	24.28%	3	3	3	9
	LibSVR	18.790	11.484	23.06%	2	2	2	6
	Linear	17.883	11.359	23.00%	1	1	1	3
3	SVR	21.810	12.224	25.69%	3	2	3	8
	LibSVR	20.688	12.266	24.24%	2	3	1	6
	Linear	20.046	12.221	24.39%	1	1	2	4

Table 3. Model Evaluation of CO Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	0.336	0.240	18.98%	2	2	2	6
	LibSVR	0.406	0.282	22.10%	3	3	3	9
	Linear	0.315	0.235	18.66%	1	1	1	3
2	SVR	0.400	0.280	22.61%	2	2	1	5
	LibSVR	0.434	0.307	24.43%	3	3	3	9
	Linear	0.371	0.279	22.86%	1	1	2	4
3	SVR	0.408	0.296	24.11%	2	2	2	6
	LibSVR	0.418	0.314	25.42%	3	3	3	9
	Linear	0.392	0.290	23.64%	1	1	1	3

Sulphur Dioxide (SO_2)

SO_2 concentration is predicted by using three models of support vector machine in regression (SVR, LibSVR, Linear). Table 4 shows the results of RMSE, MAE, and RE in predicting the SO_2 concentration by using those methods. Prediction for day 1, it reveals that linear SVR is the best proposed model since it has lowest RMSE, MAE and RE. Along with day 1, prediction of day 2 and day 3 also reveals that the linear SVR is the best proposed model due to lowest value of RMSE, MAE and RE. Nevertheless, as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE.

Table 4. Model Evaluation of SO_2 Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	0.514	0.397	13.50%	3	3	1	7
	LibSVR	0.007	0.006	274.86%	2	2	3	7
	Linear	0.002	0.001	34.49%	1	1	2	4
2	SVR	0.379	0.293	9.48%	1	2	1	4
	LibSVR	0.007	0.007	298.95%	3	1	3	7
	Linear	0.379	0.293	9.48%	1	2	1	4
3	SVR	0.413	0.319	10.34%	3	3	1	7
	LibSVR	0.007	0.006	270.99%	2	2	3	7
	Linear	0.002	0.001	41.65%	1	1	2	4

Ground-level ozone (O_3)

The performance of support vector machine in regression for predicting the O3 concentration (SVR, LibSVR, Linear) is shown in table 5. Thus, Prediction for day 1,2, and 3 shows that linear SVR is the best proposed model since it has lowest RMSE, MAE and RE. As well as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE. Moreover, SVR (SVM) and SVR (Lib SVM) is the second and third best model in predicting the O_3 concentration.

Table 5. Model Evaluation of O_3 Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	0.005	0.004	39.93%	1	1	2	4
	LibSVR	0.031	0.03	356.40%	3	3	3	9
	Linear	0.005	0.004	39.90%	1	1	1	3
2	SVR	0.006	0.005	48.59%	1	2	2	5
	LibSVR	0.02	0.019	242.60%	3	3	3	9
	Linear	0.006	0.004	44.33%	1	1	1	3
3	SVR	0.007	0.005	49.41%	2	1	2	5
	LibSVR	0.03	0.03	356.12%	3	3	3	9
	Linear	0.006	0.005	46.58%	1	1	1	3

Nitrogen Dioxide (NO_2)

NO_2 concentration is predicted by using three models of support vector machine in regression (SVR, LibSVR, Linear). Table 6 shows the results of RMSE, MAE, and RE in predicting the NO_2 concentration by using those methods. Prediction for day 1, it reveals that SVR is the best proposed model since it has lowest RMSE, MAE and RE. Along with day 1, prediction of day 2 and day 3 also reveals that the SVR is the best proposed model due to lowest value of RMSE, MAE and RE. Nevertheless, as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE.

Table 6. Model Evaluation of NO_2 Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	0.006	0.005	17.70%	1	1	1	3
	LibSVR	0.010	0.008	42.68%	3	3	3	9
	Linear	0.006	0.005	21.89%	1	1	2	4
2	SVR	0.007	0.006	21.19%	1	1	1	3
	LibSVR	0.011	0.009	47.68%	3	3	3	9
	Linear	0.007	0.006	26.68%	1	1	2	4
3	SVR	0.008	0.006	23.36%	2	1	1	4
	LibSVR	0.010	0.009	45.70%	3	3	3	9
	Linear	0.007	0.006	26.76%	1	1	2	4

Particulate Matter ($PM_{2.5}$)

The performance of support vector machine in regression for predicting the $PM_{2.5}$ concentration (SVR, LibSVR, Linear) is shown in table 7. Prediction for day 1,2, and 3 shows that linear SVR is the best proposed model since it has lowest RMSE, MAE and RE. As well as the predicted day increase, the performance of the model decreases as indicated by the increase in RMSE, MAE and RE..

Table 7. Model Evaluation of $PM_{2.5}$ Concentration Prediction

Day	Model	Model Evaluation			Rank			
		RMSE	MAE	RE	RMSE	MAE	RE	Sum
1	SVR	8.167	5.342	20.21%	1	2	2	5
	LibSVR	12.818	6.331	22.04%	3	3	3	9
	Linear	8.576	5.302	20.17%	2	1	1	4
2	SVR	9.963	6.221	24.79%	1	1	1	3
	LibSVR	12.81	6.987	26.22%	3	3	3	9
	Linear	9.963	6.221	24.79%	1	1	1	3
3	SVR	10.7	6.713	26.44%	1	1	1	3
	LibSVR	13.215	7.246	27.26%	3	3	3	9
	Linear	10.7	6.713	26.44%	1	1	1	3

Table 8. The Best Prediction Model for Each Pollutants

Day	Model	PM_{10}	CO	SO_2	O_3	NO_2	$PM_{2.5}$
1	SVR					√	
	LibSVR						
	Linear	√	√	√	√		√
2	SVR					√	
	LibSVR						
	Linear	√	√	√	√		√
3	SVR					√	
	LibSVR						
	Linear	√	√	√	√		√

SVR and LibSVR is the second and third best model in predicting the $PM_{2.5}$ concentration. As seen in table 8 the best prediction model for six major pollutants (SO_2, $PM_{2.5}$, PM_{10}, O_3, NO_2, CO).

Conclusions

This research focused on six major pollutants at Petaling Jaya station like particulate matter (PM_{10}, $PM_{2.5}$), sulphur dioxide (SO_2) nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3). This study involved secondary data of 18 years (2002-2020) obtained from Department of Environment Malaysia. Based on this study, PM_{10} has many outliers compared to others air pollutants namely $PM_{2.5}$, SO_2, NO_2, CO, and O_3 which indicates that PM_{10} has a wobbling dataset. In addition, PM_{10} has the highest standard deviation which means that the PM_{10} concentrations are far from the mean of set and spread over. Moreover, since the value of skewness tend to skew to the left and right. Therefore, the support vector regression (SVR) was used is this study.

Besides, this study aims to identify the best model in predicting the air pollutant concentration among those three types of SVR (linear, SVR, libSVR). It can be concluded that, the best model is linear SVR with the lowest total rank value of RMSE, MAE, and RE for five out of six pollutants in this study having an average of 5.548, 3.490, and 27.98% respectively. The five air pollutants include particulate matter (PM_{10}, $PM_{2.5}$), sulphur dioxide (SO_2), carbon monoxide (CO), ozone (O_3). Unlikely for nitrogen dioxide (NO_2), the best model was support vector regression (SVR) by having the average value of RMSE (0.007), MAE (0.006), and RE (20.75%) in predicting the air pollutant concentration.

Finally, the proposed model can accurately predict the daily air pollution episodes over three consecutive days and could potentially be used as an early warning system to alert local authorities about the air quality. Additionally, good air quality plays a significant role in supporting biodiversity and maintaning healthy ecosystems. However, it is only applicable in this particular location and when the source's characteristics remain constant.

Acknowledgments

The research was funded by Ministry of Science, Technology & Innovation (MOSTI) under Tech-nology Development Fund 1(TDF04211363). Thank you to School of Mathematical Sciences, College of Computing, Informatics and Media, Universiti Teknologi MARA for their support and also thanks to the Department of Environment Malaysia for providing air quality monitoring data.

References

[1] K.S. Wong, Y.J. Chew, S.Y. Ooi, Y H. Pang, *Toward forecasting future day air pollutant index in Malaysia,* **The Journal of Supercomputing**, **77**(5), 2020, pp. 4813–4830. https://doi.org/10.1007/s11227-020-03463-z.

[2] S.M. Cabaneros, J.K. Calautit, B.R. Hughes, *A review of artificial neural network models for ambient air pollution prediction*, **Environmental Modelling & Software**, **119**, 2019, pp. 285–304. https://doi.org/10.1016/j.envsoft.2019.06.014.
[3] D. Zhang, S.S. Woo, *Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network*, **IEEE Access**, **8**, 2020, pp. 89584–89594. https://doi.org/10.1109/access.2020.2993547.
[4] H. Baltaci, C.S. Ozgen Alemdar, B.O. Akkoyunlu, *Background atmospheric conditions of high PM10 concentrations in Istanbul, Turkey*, **Atmospheric Pollution Research**, **11**(9), 2020, pp. 1524–1534. https://doi.org/10.1016/j.apr.2020.06.020.
[5] C. Puspitasari, Nur Rokhman, Wahyono, Prediction of Ozone (O3) Values Using Support Vector Regression Method, **Jurnal Informatika Polinema**, **7**(4), 2021, pp. 81–88. https://doi.org/10.33795/jip.v7i4.777.
[6] A. Suárez Sánchez, P. García Nieto, P. Riesgo Fernández, J. del Coz Díaz, F. Iglesias-Rodríguez, Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain), **Mathematical and Computer Modelling**, **54**(5–6), 2011, pp. 1453–1466. https://doi.org/10.1016/j.mcm.2011.04.017.
[7] W. Yang, M. Deng, F. Xu, H. Wang, *Prediction of hourly PM2.5 using a space-time support vector regression model,* **Atmospheric Environment**, **181**, 2018, pp. 12–19. https://doi.org/10.1016/j.atmosenv.2018.03.015.
[8] A.J. Smola, B. Schölkopf, *A tutorial on support vector regression,* **Statistics and Computing, 14**(3), pp. 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.
[9] H. Karimian, Y. Chen, T. Tao, L. Yaqian, *Spatiotemporal analysis of air quality and its relationship with meteorological factors in the Yangtze River Delta,* **Journal of Elementology**, **3**, 2020. https://doi.org/10.5601/jelem.2019.24.4.1931.
[10] F.F. Sukatis, N.M. Noor, N.A. Zakaria, A.Z. Ul-Saufie, S. Anas, *Estimation of missing values in air pollution dataset by using various imputation methods,* **International Journal of Conservation Science**, **10,** 2019, pp. 791-804.
[11] T. Xu, H. Yan, Y. Bai, *Air Pollutant Analysis and AQI Prediction Based on GRA and Improved SOA-SVR by Considering COVID-19*, **Atmosphere**, **12**(3), 2021, Article Number: 336. https://doi.org/10.3390/atmos12030336.
[12] A.Z.U. Saufie, A.S. Yahaya, N.A. Ramli, H.A. Hamid, *PM10 Concentrations Short Term Prediction Using Feedforward Backpropagation and General Regression Neural Network in a Sub-urban Area*, **Journal of Environmental Science and Technology**, **8**(2), 2015, 59–73. https://doi.org/10.3923/jest.2015.59.73.
[13] V.N. Vapnik, **The Nature of Statistical Learning Theory,** Springer-Verlag New York, Inc.1995
[14] S. Rüping, **MySVM-Manual,** University of Dortmund, Lehrstuhl Informatik 8, 2000. http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/).
[15] C.C. Chang, C.J. Lin, *LIBSVM: a library for support vector machines,* **ACM Transactions on Intelligent Systems and Technology, 2**(3), 2011, pp. 1-27. http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Received: January 19, 2023
Accepted: September 12, 2023