Improving the Regression of Air Quality Using Ensemble of Machine Learning Models

Hamzah A. Alsayadi[1,*], Abdelaziz A. Abdelhamid [2], El-Sayed M. El-Kenawy [3], Abdelhameed Ibrahim [4], Marwa M. Eid [5]
[1] Computer Science Department, Faculty of Sciences, Ibb University, Yemen
[2] Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, 11566, Egypt
[3] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt
[4] Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, 35516, Mansoura Egypt
[5] Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt
Email: hamzah.sayadi@cis.asu.edu.eg, abdelaziz@cis.asu.edu.eg, skenawy@ieee.org, afai79@mans.edu.eg, marwa.3eeed@gmail.com

Abstract

Air pollution is a particularly important problem in most countries right now because of its terrible effects on both the environment and human health. Big cities are most impacted because of the country's quick industrial and economic development. In this paper, the authors proposed various regression model for the prediction of air quality including decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. The air quality dataset, in Itally cities, is used for training and evaluation the proposed model. The results show that there is a decrease in RMSE, MAE, MBE, R, R^2, RRMSE, NSE, and WI when compared to the traditional methods.

Keywords: Air quality; Ensemble model; Machine learning; Regression model

1 introduction

Air is one of the basics of life for any living creature and it is indispensable. Thus, air quality plays an important role in human health. In urban areas and cities, the air pollution is increase due to the rapid development of the world economy and the acceleration of industrialization. The air pollution has a huge impact on human live [1]. Generally, the pollution is divided into two types: (1) natural pollution because of volcanic eruptions and forest fires resulting in emission of SO2, CO2, CO, NO2, and sulfate as air pollutants and (2) man-made pollution because of some human activities such as burning of oils, discharges from industrial production processes, and transportation emissions that have PM2.5 as its major air pollutant [2] which has received much attention due to their destructive effects on human health, other kinds of creatures, and environment [3]. To make an environment suitable for human health, air pollution must be decrease. Recently, the prediction of air quality represents a big challenge due to this pollution. There are several approaches to predict air pollution such as physical prediction approaches and mathematical prediction approaches [4]. The physical prediction models are based on aerodynamics, atmospheric physics, and chemistry to study pollutant diffusion mechanism [5], and use mathematical equations to compute the spatiotemporal distributions of pollutants [6, 7]. The mathematical prediction models are based on statistics and use historical time series data to predict

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022

future air quality [4]. The mathematical prediction approaches can be imprecisely divided into three traditional classes: (1) statistical forecasting methods, (2) artificial intelligence methods, and (3) numerical forecasting methods [8]. There are a variety of machine learning algorithms in air pollution forecasting applications such as multiple linear regression model [9], Neural Networks algorithm [1], Fuzzy logic algorithm [10, 11] and Support Vector Regression [12, 13].

In this proposed approach, five different regression models which include regression models such decision tree regressor, Multi-Layer Perceptron Model (MLP) regressor, support vector regression (SVR), random forest regressor, and K-Neighbors regressor to build five model have been implemented to predict air quality. Also, the proposed have been evaluated using statistical metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R2. Results show the achievement of better performance with decreased error rate when compared to traditional prediction models.

2 Related Works

There are prediction models that presented in air pollution forecasting applications. In this section we present some work in air quality prediction.

Bai Heming [14] used the neural networks to predict forecasting PM10 values and Air Quality Index. The Markov chain stochastic process and VAR-NN-PSO were used by Caraka et al. [15] to predict PM2.5. For the datasets collected in Chaozhou and Pingtung from January 2014 to May 2019, it is possible to categorise PM2.5 according to its higher likelihood of passing through the lower respiratory tract into no risk (1–30), medium risk (30–48), and moderate risk (¿49). The work [16] reported a spatial ensemble model to predict PM2.5 for the Beijing railway station, but it is not reliable for other locations. Zhang et al. [17] presented a model to predict PM2.5. The model is trained and evaluated using the historical datasets and predictive datasets. They used different metrics such as Symmetric Mean Absolute Percentage Error (SMAPE), MAE, and RMSE.

Kim et al. [18] utilized deep learning methods to build model for predicting respiratory disease risk. They also studied the effects of the indoor PM2.5 pollutant. The artificial neural network used by Xiao et al. [19] to estimate daily average PM2.5 concentrations was improved by incorporating air mass trajectory analysis and wavelet modification.

Deters et al. [20] and Sallauddin et al. [13] compared a machine learning algorithm to BT, L-SVM [21], and ANN regression models in order to predict PM2.5 concentrations using wind direction, speed, and rainfall levels in Belisario and Cotocollao over a period of six years. The strong correlation between estimated and actual data for a time series analysis during the wet season confirms a more accurate prediction of PM2.5 when the climatic conditions are becoming more hazardous or when there are high levels of precipitation or strong winds. Beelen et al. [9] developed a multicenter cohort study in Europe in order to investigate the positive link between PM2.5 concentration and heart disease mortality over a long period of exposure to PM2.5. To predict PM2.5 time series data and other pollutants in seven locations for just 48 hours using real-time Taiwan and Beijing datasets, Soh et al. [22] acknowledged the data-driven model ST-DNN.

The performance and time complexity of the air pollutant prediction models were enhanced and the time complexity was decreased by Heni et al. [23] and Li et al. [24] using multivariate multistep time series prediction with random forest models. Zhao et al. [25] and Ni et al. [26] used a multivariate linear regression model to achieve short period prediction of PM2.5. The parameters included data on aerosol optical depth obtained through remote sensing and meteorological factors from ground monitoring temperature, relative humidity, and wind velocity. In [12], a deep learning model comprised of a recurrent neural network with long-short-term memory is used to forecast local 8-hour averaged surface ozone concentrations for 72 hours based on hourly air quality measurements. Meteorological data measurements were also used as a tool to forecast air pollution values with a lower error rate.

Ameer et al. [27] compared decision tree, random forest gradient boosting, and ANN [28] multilayer perceptron regression with respect to error rate and processing time for forecasting air quality in smart cities based

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022

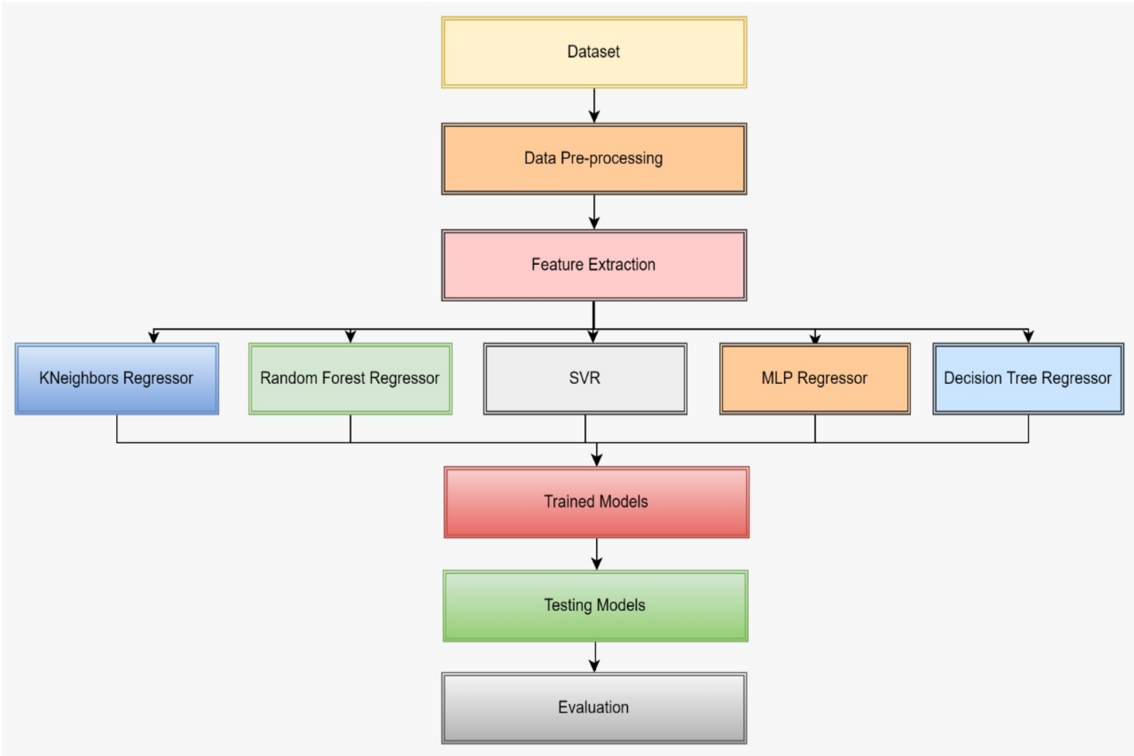

Figure 1: Regression model for the air quality prediction.

on the effects of PM2.5 during the past 25 years. In order to anticipate air quality through data exploration in three monitoring areas in Wuhan City with high accuracy, Lin et al. [11] suggested a novel approach based on a cloud model granulation technique.

3 Methodology

In this work, the regression model are used to predict the air quality. We five regression models including decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. Figure 1 represents the overview of regression model for predicting the air quality. It comprises three levels: a) the first level includes data pre-prospecting and feature extraction; b) the second level he second layer contains five different regression models which are used to predict the air quality along with its working principle; c) the last level consists of certain steps like training and testing models and evaluation, and then the final step for prediction.

3.1 Data Pre-processing and Feature extraction

The data were preprocessed to remove the missing and unwanted data to obtain a cleaned datasets. Then, the features are extracted from the cleaned data to obtain the best features for training models.

3.2 The proposed models

We proposed decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor to build five model. We give a brief overview for these method as follows.

Decision Trees. In order to classify or predict data, decision trees use a tree-like model of decisions and their outcomes. A classification tree and prediction both benefit from the use of a decision tree (regression

https://doi.org/10.54216/JAIM.010201

Received: January 20, 2022 Accepted: May 25, 2022

tree). Because it is less reliable than a random forest regression, a single regression tree may not be able to forecast complex situations (RFR). RFR is the most often used decision tree-based model as it is an ensemble of regression trees [29, 30, 31].

Multi-Layer Perceptron Model. The Multilayer Perceptron (MLP) is a feed forward neural network type that is used for predicting air pollution because of its capacity to create highly complex nonlinear models. The MLP begins when the network receives the interested input parameters. These input parameters produce input signals, which are transmitted across the network starting from the input layer to the hidden layer and then the hidden layer to the output layer. The weights, which are a real numerical amount, are multiplied by the scaled input vector that the input layer's neurons introduce [32, 33, 34].

Support Vector Regression. One of the uses of support vector machines (SVM), which were initially introduced in 1954, is support vector regression (SVR). The fundamental concepts of SVM are structural minimization and statistical learning theory. The primary goal of the kernel function is to convert data from a low-dimensional to a high-dimensional space, which transforms data from a nonlinear to a linear feature space. Radial basis functions, linear basis functions, and polynomial basis functions are the typical kernel functions that can be employed in SVR [35, 36, 37].

Random Forest. An ensemble learning-based machine learning technique called random forest (RF) can be utilized for both classification and regression problems. The supervised machine learning algorithms category includes it. In order to create the random forest architecture, a group of trees are built in parallel using the bagging technique, which is the fundamental approach utilized in random forests. The random forest's trees are constructed, but there is no interaction between them. This method's training phase involves building a large number of decision trees that may be applied to either classification or regression. The result of random forest in regression problems is the average forecast made by each tree. Therefore, the prediction using random forest is considered a combination of the multiple predictions achieved by the constructed decision trees [38**?**].

k Nearest Neighbors. One of the most popular machine learning algorithms is the k Nearest Neighbors (kNN) technique. It has categorization applications. Machine learning uses statistical and mathematical techniques to draw conclusions from available data. An algorithm for non-parametric learning is kNN. The k nearest neighbors algorithm groups new cases according to distance function criteria and saves all of the current cases. The data used for training is memorized rather than learned. The algorithm's operation results in the determination of a k value. The number of components to consider is k. The distance between the value and the closest k element is determined when a value is reached. The instance is merely given to its closest neighbor if k is 1. The best k numbers historically ranged from 3 to 10 for the majority of data sets. For calculating distances, the Euclidean function is typically used. Alternatives to the Euclidean function include the Manhattan, Minkowski, and Hamming functions [39].

The Proposed Ensemble Model. After optimizing the parameters of each regression model in the proposed ensemble, these optimized models are integrated into a unified ensemble. This ensemble consists of five regression models including Decision Tree Regressor, MLP Regressor, SVR, Random Forest Regressor, and KNeighbors Regressor. The output of the ensemble model is the optimal predicted value that best matches the input features of the air quality.

3.3 Regression Model Training

After feature extraction step, the features represent the data training for models. Five separate models were subsequently created for testing and training. In order to deploy the models, you must first confirm the prediction value range. If necessary, you should then forecast if the air quality levels are satisfactory or good; if not, the models and datasets need to be improved once more.

3.4 Evaluation

The most popular evaluation criteria are the correlation coefficient (R2), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Relative Absolute Error (RAE). The R2 value shows the fitting degree of

Received: January 20, 2022 Accepted: May 25, 2022

Table 1: The results of the regression model

Model	RMSE	MAE	MBE	R	R^2	RRMSE	NSE	WI
Decision Tree	0.072	0.055	-0.0099	0.9491	0.9008	18.208	0.8988	0.8612
Multi-layer Perceptron	0.055	0.0408	-0.0074	0.9713	0.9435	13.8316	0.9416	0.8973
support vector regression	0.0489	0.03904	0.0095	0.9776	0.9558	12.29	0.9538	0.9019
Random forest	0.1051	0.0805	0.0008	0.8878	0.7882	26.43	0.7867	0.7976
K-nearest neighbors	0.0470	0.0347	0.0041	0.9787	0.9578	11.81	0.9573	0.9127

regression, MAE represents the difference between predicted and actual values, RMSE focuses on the impact of extreme values based on MAE, while RAE calculates the variance of a model when comparing the performance of different models. As MAE and RMSE depend on the scale of the data that's why RAE can be extremely helpful when comparing different data with different scales. The R2, MAE, RMSE and RAE are calculated by using Eqs. 1 to 4, respectively.

$$R^2 = (\frac{\sum_i [(x_i - \overline{x})(y_i - \overline{y})]}{[(\sum_i x_i - \overline{x})^2 (\sum_i y_i - \overline{y})^2]]^{\frac{1}{2}}})^2 \tag{1}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i| \tag{2}$$

$$RMSE = [\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2]^{\frac{1}{2}} \tag{3}$$

4 Experimental Results

4.1 Dataset

The dataset includes 9358 examples of hourly averaged responses from a group of five metal oxide chemical sensors that are built into an Air Quality Chemical Multisensor Device. The device was situated on a field at road level in a heavily polluted area of an Italian city. Data were collected from March 2004 to February 2005 (a period of one year), making it the longest free recording of responses from on-field installed chemical air quality sensor devices. A co-located reference certified analyzer supplied Ground Truth hourly averaged readings for CO, Non Metanic Hydrocarbons, Benzene, Total Nitrogen Oxides (NOx), and Nitrogen Dioxide (NO2).

4.2 Results and discussion

We evaluate the proposed models based on the air quality dataset of polluted area in Italian city. The metrics, that are presented in subsection 3.4, are used to evaluated the proposed model. The results of the proposed model been shown in Table 1.

As shown in Table 1, we note that the experimental results enhance the results. It is further observed all the five models produce good results.

In terms of RMSE for the proposed models, the results that are obtained 0.0724, 0.0550, 0.0489, 0.1051, and 0.0470 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022

Table 2: The results of the ensemble model

Model	**RMSE**	**MAE**	**MBE**	**R**	**R^2**	**RRMSE**	**NSE**	**WI**
Average Ensemble	0.0535	0.0399	-0.0005	0.9735	0.9478	11.81	0.9446	0.8997
Ensemble using KNN regressor	0.0353	0.03339	0.0031	0.9801	0.9606	11.38	0.9604	0.9161

In terms of MAE for the proposed models, the results that are obtained 0.0552, 0.0408, 0.0390, 0.0805, and 0.0347 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

In terms of MBE for the proposed models, the results that are obtained -0.0099, -0.0074, 0.0095, 0.0008, and 0.0041 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the random forest regressor model achieved the best results.

In terms of R for the proposed models, the results that are obtained 0.949108996220673, 0.9713, 0.9776, 0.8878, and 0.9787 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

In terms of R2 for the proposed models, the results that are obtained 0.9008078867070135, 0.9435, 0.9558, 0.7882, and 0.9578 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

In terms of RRMSE for the proposed models, the results that are obtained 18.20, 13.83, 12.29, 26.43, and 11.81 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

In terms of NSE for the proposed models, the results that are obtained 0.8988, 0.9416, 0.9538, 0.7867, and 0.9573 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the decision tree regressor model achieved the best results.

In terms of WI for the proposed models, the results that are obtained 0.8612, 0.8973, 0.9019, 0.7976, and 0.9127 for decision tree regressor, MLP regressor, SVR, random forest regressor, and KNeighbors regressor, respectively. We observe that the KNeighbors regressor model achieved the best results.

Two experiments of ensembles are conducted in this work including average ensemble and ensemble using K-Neighbors Regressor. These models are evaluated based the RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI metrics as shown in Table 2.

The average ensemble yielded the results of RMSE, MAE, MBE, R, R2, RRMSE, NSE, and WI by 0.0535, 0.0399, -0.0005, 0.9735, 0.9478, 11.81, 0.9446, and 0.8997, respectively. While ensemble using K-Neighbors regressor achieved the results of RMSE, MAE, MBE, R, R^2, RRMSE, NSE, and WI by 0.0353, 0.0333, 0.0031, 0.9801, 0.9606, 11.38, 0.9604, and 0.9161.

5 Conclusions

In this paper, we proposed the regression models to predict the air quality including decision tree regressor, MLP regressor, SVR, random forest regressor, and K-Neighbors regressor. The average ensemble and ensemble using K-Neighbors Regressor are used for evaluation the proposed models. We used MSE, MAE, MBE, R, R2, RRMSE, NSE, and WI metrics to evaluate the proposed regression models and ensemble models. The KNeighbors regressor model achieved the best results for RMSE, MAE, R, $R^@$, RRMSE, and WI. While the random forest regressor model achieved the best results for MBE. The decision tree regressor model achieved the best results for NSE.

Received: January 20, 2022 Accepted: May 25, 2022

References

[1] Bing-Chun Liu, Arihant Binaykia, Pei-Chann Chang, Manoj Kumar Tiwari, and Cheng-Chin Tsao. Urban air quality forecasting based on multi-dimensional collaborative support vector regression (svr): A case study of beijing-tianjin-shijiazhuang. *PloS one*, 12(7):e0179763, 2017.

[2] Lu Bai, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. Air pollution forecasts: An overview. *International journal of environmental research and public health*, 15(4):780, 2018.

[3] Andrew C Kemp, Benjamin P Horton, Jeffrey P Donnelly, Michael E Mann, Martin Vermeer, and Stefan Rahmstorf. Climate related sea-level variations over the past two millennia. *Proceedings of the National Academy of Sciences*, 108(27):11017–11022, 2011.

[4] Wenjing Mao, Weilin Wang, Limin Jiao, Suli Zhao, and Anbao Liu. Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustainable Cities and Society*, 65:102567, 2021.

[5] Guannan Geng, Qiang Zhang, Randall V Martin, Aaron van Donkelaar, Hong Huo, Huizheng Che, Jintai Lin, and Kebin He. Estimating long-term pm2. 5 concentrations in china using satellite-based aerosol optical depth and a chemical transport model. *Remote sensing of Environment*, 166:262–270, 2015.

[6] Jaein I Jeong, Rokjin J Park, Jung-Hun Woo, Young-Ji Han, and Seung-Muk Yi. Source contributions to carbonaceous aerosol concentrations in korea. *Atmospheric environment*, 45(5):1116–1125, 2011.

[7] Yunhee Kim, Joshua S Fu, and Terry L Miller. Improving ozone modeling in complex terrain at a fine grid resolution: Part i–examination of analysis nudging and all pbl schemes associated with lsms in meteorological model. *Atmospheric Environment*, 44(4):523–532, 2010.

[8] D Kothandaraman, N Praveena, K Varadarajkumar, B Madhav Rao, Dharmesh Dhabliya, Shivaprasad Satla, and Worku Abera. Intelligent forecasting of air quality and pollution prediction using machine learning. *Adsorption Science & Technology*, 2022, 2022.

[9] Rob Beelen, Ole Raaschou-Nielsen, Massimo Stafoggia, Zorana Jovanovic Andersen, Gudrun Weinmayr, Barbara Hoffmann, Kathrin Wolf, Evangelia Samoli, Paul Fischer, Mark Nieuwenhuijsen, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 european cohorts within the multicentre escape project. *The lancet*, 383(9919):785–795, 2014.

[10] SN Pasha, A Harshavardhan, D Ramesh, S Shabana Md, et al. Variation analysis of artificial intelligence machine learning and advantages of deep architectures. *International Journal of Advanced Science and Technology*, 28(17):488–495, 2019.

[11] Yi Lin, Long Zhao, Haiyan Li, and Yu Sun. Air quality forecasting based on cloud model granulation. *EURASIP Journal on Wireless Communications and Networking*, 2018(1):1–10, 2018.

[12] Brian S Freeman, Graham Taylor, Bahram Gharabaghi, and Jesse Thé. Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, 68(8):866–886, 2018.

[13] M Sallauddin, D Ramesh, A Harshavardhan, SN Pasha, and A Shabana. A comprehensive study on traditional ai and ann architecture. *International Journal of Advanced Science and Technology*, 28(17):479–487, 2019.

[14] He-Ming Bai, Run-Ping Shen, Hua-Ding Shi, and Yu-Chun Dong. Forecasting model of air pollution index based on bp neural network. *Huanjing Kexue yu Jishu*, 36(3):186–189, 2013.

[15] Rezzy Eko Caraka, Rung Ching Chen, Toni Toharudin, Bens Pardamean, Hasbi Yasin, and Shih Hung Wu. Prediction of status particulate matter 2.5 using state markov chain stochastic process and hybrid var-nn-pso. *IEEE Access*, 7:161654–161665, 2019.

[16] Yinan Xu and Hui Liu. Spatial ensemble prediction of hourly pm2. 5 concentrations around beijing railway station in china. *Air Quality, Atmosphere & Health*, 13(5):563–573, 2020.

[17] Ying Zhang, Yanhao Wang, Minghe Gao, Qunfei Ma, Jing Zhao, Rongrong Zhang, Qingqing Wang, and Linyan Huang. A predictive data feature exploration-based air quality prediction approach. *IEEE Access*, 7:30732–30743, 2019.

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022

[18] Dohyeong Kim, Sunghwan Cho, Lakshman Tamil, Dae Jin Song, and Sungchul Seo. Predicting asthma attacks: effects of indoor pm concentrations on peak expiratory flow rates of asthmatic children. *IEEE Access*, 8:8791–8797, 2019.

[19] Xiao Feng, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, and Jingjie Wang. Artificial neural networks forecasting of pm2. 5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*, 107:118–128, 2015.

[20] Jan Kleine Deters, Rasa Zalakeviciute, Mario Gonzalez, and Yves Rybarczyk. Modeling pm2. 5 urban pollution using machine learning and selected meteorological parameters. *Journal of Electrical and Computer Engineering*, 2017, 2017.

[21] A Harshavardhan, Dr Suresh Babu, and Dr T Venugopal. An improved brain tumor segmentation and classification method using svm with various kernels. *J. Int. Pharma. Res*, 46(2):489–495, 2019.

[22] Ping-Wei Soh, Jia-Wei Chang, and Jen-Wei Huang. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *Ieee Access*, 6:38186–38199, 2018.

[23] Heni Patel and Swarndeep Saket. Air pollution prediction system for smart city using data mining technique: A survey. *health*, 6(12), 2019.

[24] Jihan Li, Xiaoli Li, and Kang Wang. Atmospheric pm2. 5 concentration prediction based on time series and interactive multiple model approach. *Advances in Meteorology*, 2019, 2019.

[25] Rui Zhao, Xinxin Gu, Bing Xue, Jianqiang Zhang, and Wanxia Ren. Short period pm2. 5 prediction based on multivariate linear regression model. *PloS one*, 13(7):e0201011, 2018.

[26] XY Ni, Hong Huang, and WP Du. Relevance analysis and short-term prediction of pm2. 5 concentrations in beijing based on multi-source data. *Atmospheric environment*, 150:146–161, 2017.

[27] Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, Saif Ul Islam, and Muhammad Nabeel Asghar. Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access*, 7:128325–128338, 2019.

[28] Venkataramana Veeramsetty and Ram Deshmukh. Electric power load forecasting on a 33/11 kv substation using artificial neural networks. *SN Applied Sciences*, 2(5):1–10, 2020.

[29] Wenjuan Wei, Olivier Ramalho, Laeticia Malingre, Sutharsini Sivanantham, John C Little, and Corinne Mandin. Machine learning and statistical models for predicting indoor air quality. *Indoor Air*, 29(5):704–726, 2019.

[30] El-Sayed M. El-Kenawy, Seyedali Mirjalili, Fawaz Alassery, Yu-Dong Zhang, Marwa Metwally Eid, Shady Y. El-Mashad, Bandar Abdullah Aloyaydi, Abdelhameed Ibrahim, and Abdelaziz A. Abdelhamid. Novel Meta-Heuristic Algorithm for Feature Selection, Unconstrained Functions and Engineering Problems. *IEEE Access*, 10:40536–40555, 2022.

[31] Abdelaziz A. Abdelhamid, El-Sayed M. El-Kenawy, Bandar Alotaibi, Ghada M. Amer, Mahmoud Y. Abdelkader, Abdelhameed Ibrahim, and Marwa Metwally Eid. Robust Speech Emotion Recognition Using CNN+LSTM Based on Stochastic Fractal Search Optimization Algorithm. *IEEE Access*, 10:49265–49284, 2022.

[32] S Abdullah, M Ismail, and AN Ahmed. Multi-layer perceptron model for air quality prediction. *Malaysian Journal of Mathematical Sciences*, 13:85–95, 2019.

[33] Doaa Sami Khafaga, Amel Ali Alhussan, El-Sayed M. El-kenawy, Ali E. Takieldeen, Tarek M. Hassan, Ehab A. Hegazy, Elsayed Abdel Fattah Eid, Abdelhameed Ibrahim, and Abdelaziz A. Abdelhamid. Metaheuristics for Feature Selection and Classification in Diagnostic Breast-Cancer. *Computers, Materials & Continua*, 73(1):749–765, 2022.

[34] Doaa Sami Khafaga, Amel Ali Alhussan, El-Sayed M. El-kenawy, Abdelhameed Ibrahim, Said H. Abd Elkhalik, Shady Y. El-Mashad, and Abdelaziz A. Abdelhamid. Improved Prediction of Metamaterial Antenna Bandwidth Using Adaptive Optimization of LSTM. *Computers, Materials & Continua*, 73(1):865–881, 2022.

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022

[35] Ruizhi Zhong, Raymond L Johnson Jr, and Zhongwei Chen. Using machine learning methods to identify coals from drilling and logging-while-drilling lwd data. In *Asia Pacific Unconventional Resources Technology Conference, Brisbane, Australia, 18-19 November 2019*, pages 970–994. Unconventional Resources Technology Conference, 2020.

[36] Nagwan Abdel Samee, El-Sayed M. El-Kenawy, Ghada Atteia, Mona M. Jamjoom, Abdelhameed Ibrahim, Abdelaziz A. Abdelhamid, Noha E. El-Attar, Tarek Gaber, Adam Slowik, and Mahmoud Y. Shams. Metaheuristic Optimization Through Deep Learning Classification of COVID-19 in Chest X-Ray Images. *Computers, Materials & Continua*, 73(2):4193–4210, 2022.

[37] Hussah Nasser AlEisa, El-Sayed M. El-kenawy, Amel Ali Alhussan, Mohamed Saber, Abdelaziz A. Abdelhamid, and Doaa Sami Khafaga. Transfer Learning for Chest X-rays Diagnosis Using Dipper Throated Algorithm. *Computers, Materials & Continua*, 73(2):2371–2387, 2022.

[38] Zhi-Hua Zhou. *Machine learning*. Springer Nature, 2021.

[39] Burhan BARAN. Air quality index prediction in besiktas district by artificial neural networks and k nearest neighbors. *Mühendislik Bilimleri ve Tasarım Dergisi*, 9(1):52–63, 2021.

https://doi.org/10.54216/JAIM.010201
Received: January 20, 2022 Accepted: May 25, 2022