

S. Sankar Ganesh\*, Pachaiyappan Arulmozhivarman and Rao Tatavarti

# Forecasting Air Quality Index Using an Ensemble of Artificial Neural Networks and Regression Models

<https://doi.org/10.1515/jisys-2017-0277>

Received June 8, 2017; previously published online November 18, 2017.

**Abstract:** Air is the most essential constituent for the sustenance of life on earth. The air we inhale has a tremendous impact on our health and well-being. Hence, it is always advisable to monitor the quality of air in our environment. To forecast the air quality index (AQI), artificial neural networks (ANNs) trained with conjugate gradient descent (CGD), such as multilayer perceptron (MLP), cascade forward neural network, Elman neural network, radial basis function (RBF) neural network, and nonlinear autoregressive model with exogenous input (NARX) along with regression models such as multiple linear regression (MLR) consisting of batch gradient descent (BGD), stochastic gradient descent (SGD), mini-BGD (MBGD) and CGD algorithms, and support vector regression (SVR), are implemented. In these models, the AQI is the dependent variable and the concentrations of  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{PM}_{2.5}$ ,  $\text{SO}_2$ , and  $\text{PM}_{10}$  for the years 2010–2016 in Houston and Los Angeles are the independent variables. For the final forecast, several ensemble models of individual neural network predictors and individual regression predictors are presented. This proposed approach performs with the highest efficiency in terms of forecasting air quality index.

**Keywords:** Artificial neural networks, air quality index, ensemble of predictors, forecasting, gradient descent.

## 1 Introduction

The quality of air significantly contributes to the health of the inhabitants in a particular area. Inhaling polluted air can cause dreadful diseases such as lung cancer, stroke, and respiratory infections, especially in children. The long-term consequences of polluted air lead to global warming and greenhouse [15]. It is a major concern in most of the densely populated areas. New diseases are being diagnosed every day and more deaths are being reported every year due to air pollution. There are several reasons for the increase in the levels of air pollution, such as industrialization and globalization [8]. A number of monitoring sites have been set up globally to monitor the quality of air [4, 18]. Air quality monitoring using soft computing techniques has produced a tremendous research analysis in recent years.

All over the world, there are more deaths due to poor air quality compared to other sources [7]. The World Health Organization (WHO) states that there are about 2.4 million deaths that are directly attributed to air pollution [19]. The concept of the air quality index (AQI) has been proposed to address the purpose of measurement of air quality in an area. Several methods under artificial intelligence (AI) domain have been proposed for forecasting. In particular, computational intelligence methods, such as artificial neural networks (ANN), contribute a significant amount of accuracy in the prediction of air quality. These methods have been proven to be a success in various scenarios. The prescribed standards of the AQI in Houston and New York are given by the U.S. Environmental Protection Agency (EPA) (<https://www3.epa.gov/airquality/cleanair.html>) and are shown in Table 1.

---

\*Corresponding author: S. Sankar Ganesh, VIT University, Vellore, India, e-mail: s.sankarganesh@vit.ac.in

Pachaiyappan Arulmozhivarman: VIT University, Vellore, India

Rao Tatavarti: GVP-SIRC, Visakhapatnam, India

**Table 1:** Proposed Categories of AQI (0–500) by the US EPA.

Range	AQI category
0–50	Good
51–100	Moderate
101–150	Unhealthy for sensitive groups
201–300	Unhealthy
301–400	Very unhealthy
401–500	Hazardous

## 2 Materials and Methods

### 2.1 Data Preparation

Houston is the most populous city in Texas State, with a population count of about 2 million (<https://en.wikipedia.org/wiki/Houston>), and Los Angeles is the second most densely populated city ([https://en.wikipedia.org/wiki/Los\\_Angeles](https://en.wikipedia.org/wiki/Los_Angeles)) in the United States, with an estimated population of 4.04 million [11, 16]. The data related to the quality of air in Houston, including  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{PM}_{2.5}$ , and  $\text{SO}_2$ , and Los Angeles, including  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$ , for the years 2010–2016 were obtained from the US EPA. The Houston and Los Angeles data containing 2000 samples were divided into 1500 samples (75%) and 500 samples (25%) for training and testing each predictive model, respectively. The proposed approach uses the cascade forward neural network ensemble of individual neural networks and regression models and also the support vector regression (SVR) ensemble of individual neural networks and regression models. The base learners include neural networks such as multilayer perceptron (MLP), cascade forward neural network, Elman neural network, radial basis function (RBF) neural network, and nonlinear autoregressive model with exogenous input (NARX) and regression models such as multiple linear regression (MLR) with gradient descent variants as optimization algorithms and SVR. In the case of Houston, for each individual neural network predictor and regression model, the AQI is the dependent variable and the data related to the concentrations of  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{SO}_2$ , and  $\text{PM}_{2.5}$  are the independent variables. The statistical measures of all data samples of Houston used in the current study are shown in Table 2. In the case of Los Angeles, for each individual neural network predictor and regression model, the AQI is the dependent variable and the data related to the concentrations of  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{PM}_{2.5}$ , and  $\text{PM}_{10}$  are the independent variables. The statistical measures of all data samples of Los Angeles used in the current study are shown in Table 3. The performance of each individual base learner and ensembles is evaluated using the error indexes such as mean absolute error (MAE), mean absolute percent error (MAPE), correlation coefficient (R), root mean square error (RMSE), and index of agreement (IA). The performance of the testing data is evaluated, as it represents the accuracy of each predictor.

**Table 2:** Statistical Measures of Houston Data Samples.

Variable	Unit	Range	Mean	SD
$\text{NO}_2$	ppb	6.36–132.71	32.70	13.95
$\text{CO}$	ppm	0.088–2.464	0.53	0.277
$\text{O}_3$	ppm	0.009–0.121	0.04	0.01
$\text{SO}_2$	ppb	0.009–0.121	6.86	8.98
$\text{PM}_{2.5}$	$\mu\text{g}/\text{m}^3$	3.84–45.65	13.00	5.00
AQI	NA	18–217	60.33	27

**Table 3:** Statistical Measures of Los Angeles Data Samples.

Variable	Unit	Range	Mean	SD
NO <sub>2</sub>	ppb	12.72–132.71	44.51	14.75
CO	ppm	0.264–4.7	1.14	0.65
O <sub>3</sub>	ppm	0.019–0.122	0.05	0.01
PM <sub>2.5</sub>	μg/m <sup>3</sup>	4.56–94.23	21.27	9.10
PM <sub>10</sub>	μg/m <sup>3</sup>	6.48–131.77	37.17	13.58
AQI	NA	35–218	89.57	35.55

## 2.2 Gradient Descent Variants: Conjugate Gradient Descent (CGD)

The CGD optimization algorithm [12] uses the line search method in the conjugate directions to achieve faster convergence. In this method, the step size or the learning rate is calculated after each iteration, which is known as the adaptive learning rate. The steps involved in the CGD optimization are given as follows:

Say for a quadratic test function,  $\phi(x) = \frac{1}{2}x^T Ax - x^T b$ , where  $A$  is assumed to be a symmetric positive definite (SPD) matrix. Minimizing this quadratic test function is equivalent to solving  $Ax = b$ .

Compute  $r_0 = Ax_0 - b$  and set  $p_0 = -r_0$ .

For  $k = 0, 1, 2, \dots$  until convergence

$$\eta_k = \frac{r_k^T r_k}{p_k^T A p_k} \quad (1)$$

$$x_{k+1} = x_k + \eta_k p_k \quad (2)$$

$$r_{k+1} = r_k + \eta_k A p_k \quad (3)$$

$$\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} \quad (4)$$

$$p_{k+1} = -r_{k+1} + \beta_k r_k \quad (5)$$

End

The first step in this method involves calculating the initial residual  $r_0$  or initial conjugate gradient direction equal to the steepest descent direction, which is the negative gradient of the function. The formula (1) gives the calculation of the optimal step length, which represents the distance that needs to be taken along the conjugate direction until it no longer descends. Using formula (2), we update the solution by adding the step to minimize the quadratic function  $\phi$  along  $x_k + \eta_k p_k$ . It is obvious that this leads to search directions that are orthogonal to each other. After this, the residual is updated according to formula (3). The conjugate direction method has a property in which the new conjugate vector  $p_{k+1}$  can be computed using the previous vector  $p_k$  using formula (4). The new direction or new step  $p_k$  will be the linear combination of the negative residual  $-r_{k+1}$  and the previous search vector  $p_k$ . The ratio of norm squared current residual to the norm squared of the previous residual gives the value of the constant  $\beta_k$  using formula (5).

## 2.3 Neural Networks

ANNs are the most widely used machine learning algorithms for time series prediction [22]. Neurons are the building blocks of these predictors. The most basic ANN is MLP. It belongs to the class of feed-forward

networks, which maps the input to an appropriate output and is commonly known for its robust universal approximation ability [13]. A basic MLP structure consists of an input layer, hidden layer, and an output layer in which both input and output layers are called visible layers. Cascade forward neural network also belongs to the class of feed-forward networks with a weight connection from the input layer to each layer in the network. It is based on the principle that feed-forward networks with a more number of layers in the architecture would learn more complex input-output relationship quickly. Elman neural network is a recurrent neural network with an extra layer called the context layer between the input and hidden layers to feed back the states of hidden units into hidden units during the next stage of input. It is mostly implemented to learn the time-varying patterns. RBF neural networks [3] are a kind of neural networks in which the radial basis kernel, such as Gaussian kernel, takes the role of the activation function. A clustering algorithm such as K-means clustering [14] is used to transform the input vector into K number of clusters. The optimal value of “K” is determined by the elbow method. NARX relates the current value of a time series with past values of the same series. The model employed in the current study has two tap delays, which means both previous and before previous input and output values are feed back to the input layer. In the entire neural network predictors implemented here, except for RBF neural network, the activation function used at each neuron is a logistic function given by Eq. (6):

$$f(x) = \frac{1}{1 + \exp(-x)}. \quad (6)$$

The errors cost function used in each of the neural network predictors is given by Eq. (7):

$$J = \frac{1}{2m} \sum_{i=1}^m (y(x^{(i)}) - t^{(i)})^2 \quad (7)$$

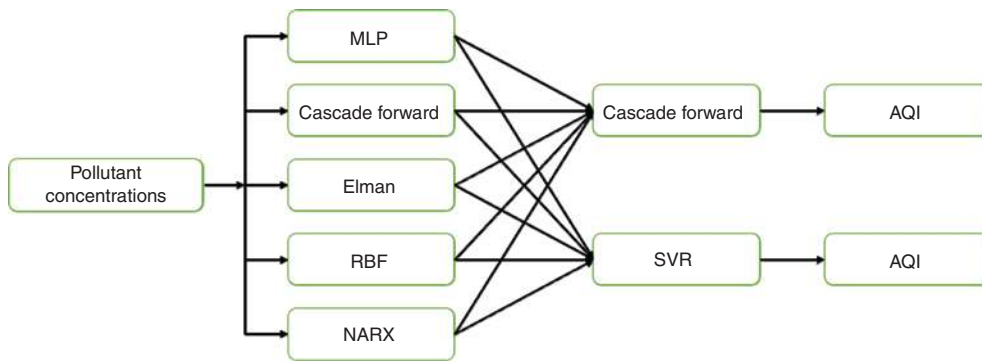
where the variables  $t^{(i)}$  and  $y(x^{(i)})$  represent the observed and predicted values of  $PM_{2.5}$ , respectively, for the  $i^{\text{th}}$  sample of input  $x$ , and  $m$  represents the number of input data samples. The neural networks implemented here relate the input and output in independent approaches. All neural networks presented in the current study have been trained using back-propagation with CGD optimization [12] to update the model parameters or weights of the network. The structural complexity of MLP, cascade forward, and NARX neural network predictors includes five input layer neurons, five hidden layer neurons, and one output layer neuron. The Elman recurrent neural network implemented in the current study has a structure of 5-5-5-1, which includes five neurons in the context layer. In the RBF neural network, the input data set has been divided into 15 clusters or it can be said that the neural network has 10 RBF neurons. These individual neural network predictors are then combined using an ensemble network consisting of cascade forward neural network with the same structured implemented before and SVR as the integrators to improve the accuracy of forecasting of the AQI. All implemented neural network predictors have used 1000 epochs. The results of both ensemble approaches are evaluated and compared.

## 2.4 Regression Models: MLR with Gradient Descent Variants and SVR

In MLR, the dependent and independent variables are related by determining the model parameters [21], and the relationship is governed by Eq. (8):

$$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad (8)$$

where  $y$  is the dependent variable,  $x_1, x_2, \dots, x_n$  are the independent variables, and  $w_1, w_2, \dots, w_n$  are the model parameters. This basic regression model is widely used in many studies for prediction [6, 10]. In the current study, this model has been implemented using batch gradient descent (BGD) with a learning rate or step size of 0.01, stochastic gradient descent (SGD) with a learning rate of 0.0001, mini-BGD (MBGD) with a learning



**Figure 1:** Cascade Forward and SVR Ensemble of Neural Network Predictors.

rate of 0.005, and batches of 50 samples and CGD optimization algorithms, as these are the most common ways of optimizing any regression model. All MLR models employed here use 1000 epochs for training. SVR [9] based on support vector machines (SVM) [5] has also been implemented as a regression model in the current study for predicting the AQI in Houston and Los Angeles.

## 2.5 Modeling Technique: Stacking Ensemble

In the current study, the stacking ensemble method is used for the final forecast of the AQI. In the stacking ensemble method, the predictions of the individual base learners are given as input to the any of the base learner algorithm, which is used as a combiner algorithm for the higher level of training. Usually, this combiner algorithm outperforms the performance of the individual base learners [20]. It has been proven to be a successful approach for supervised learning tasks such as regression analysis [2] and also for unsupervised learning tasks such as density estimation [17]. In general, a simple logistic regression is used as a combiner algorithm. The combiner algorithms used here include cascade forward neural network and SVR. All other predictors including these two are considered as base learners. Here, the performance of both combiner algorithms have been evaluated by taking two different sets of base learners: the first set of base learners includes all neural network predictors such as MLP, cascade forward neural network, Elman recurrent neural network, RBF neural network, and NARX neural network with CGD as the optimization algorithm to update weights, and the second set of base learners includes MLR with different gradient descent variants such as BGD, SGD, MBGD, and CGD optimization algorithms and SVR. The predictions of these individual neural network predictors and regression models are given as input to the combiner algorithm such as cascade forward neural network and SVR. The final forecast of this combiner algorithm predicts the AQI in an area of interest. This stacked model often improves the accuracy due to its smoothing nature and its ability to credit each of the base learners where they perform best and to discredit where they perform poorly. The overall performance of this stacked model will be good when the base learners are significantly different in the way they predict. The block diagrams demonstrating ensembles are given in Figures 1 and 2.

## 3 Results and Discussion

### 3.1 Cascade Forward and SVR Ensembles of Neural Network Predictors

In these two ensembles, the base learners are individual neural network predictors: MLP, cascade forward neural network, Elman neural network, RBF neural network, and NARX. The predictions of these predictors

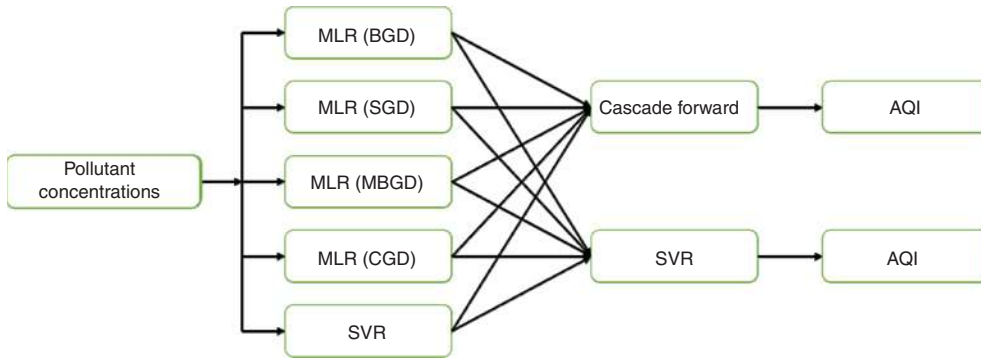


Figure 2: Cascade Forward and SVR Ensemble of MLR with Gradient Descent Variants and SVR.

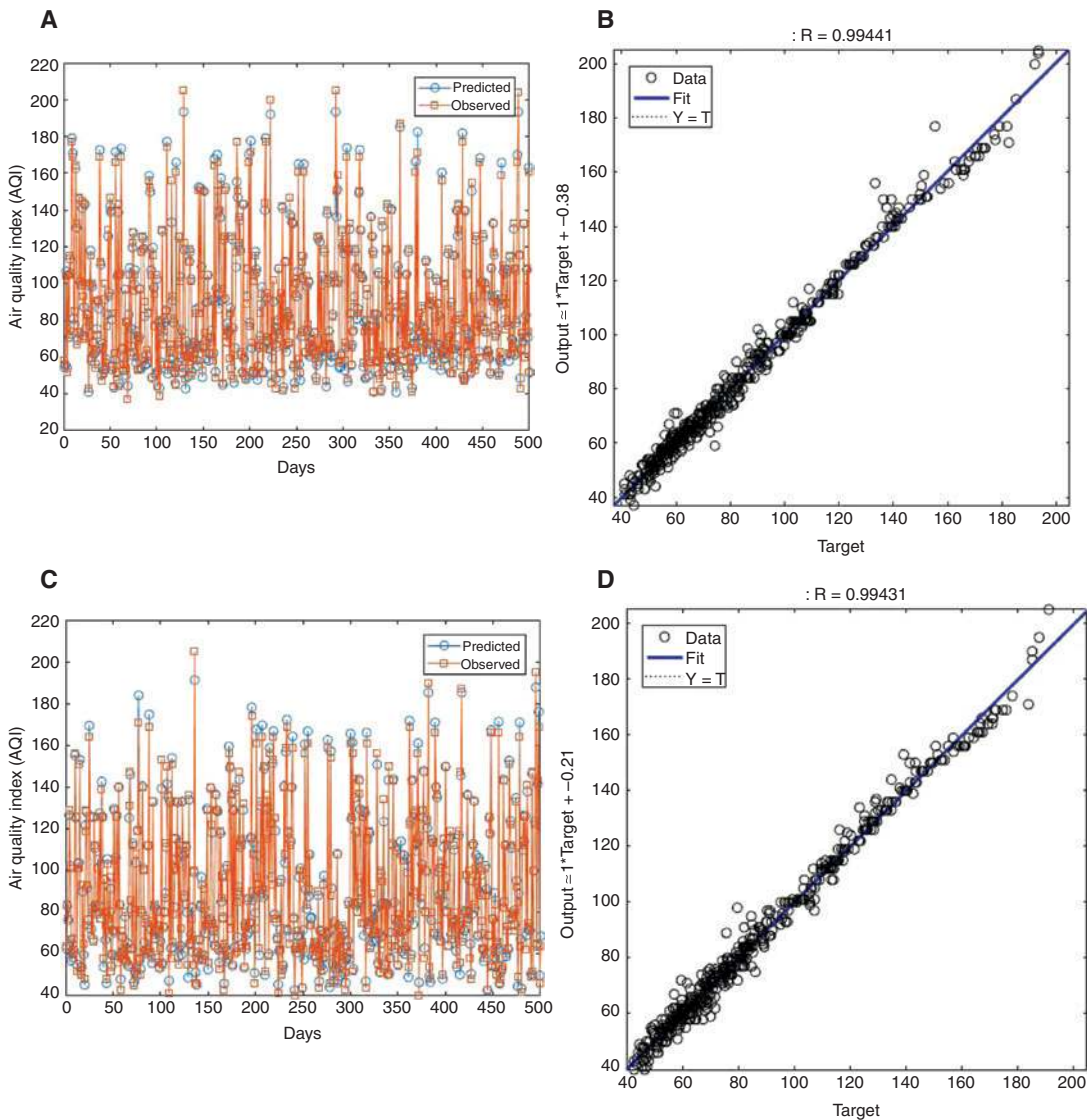


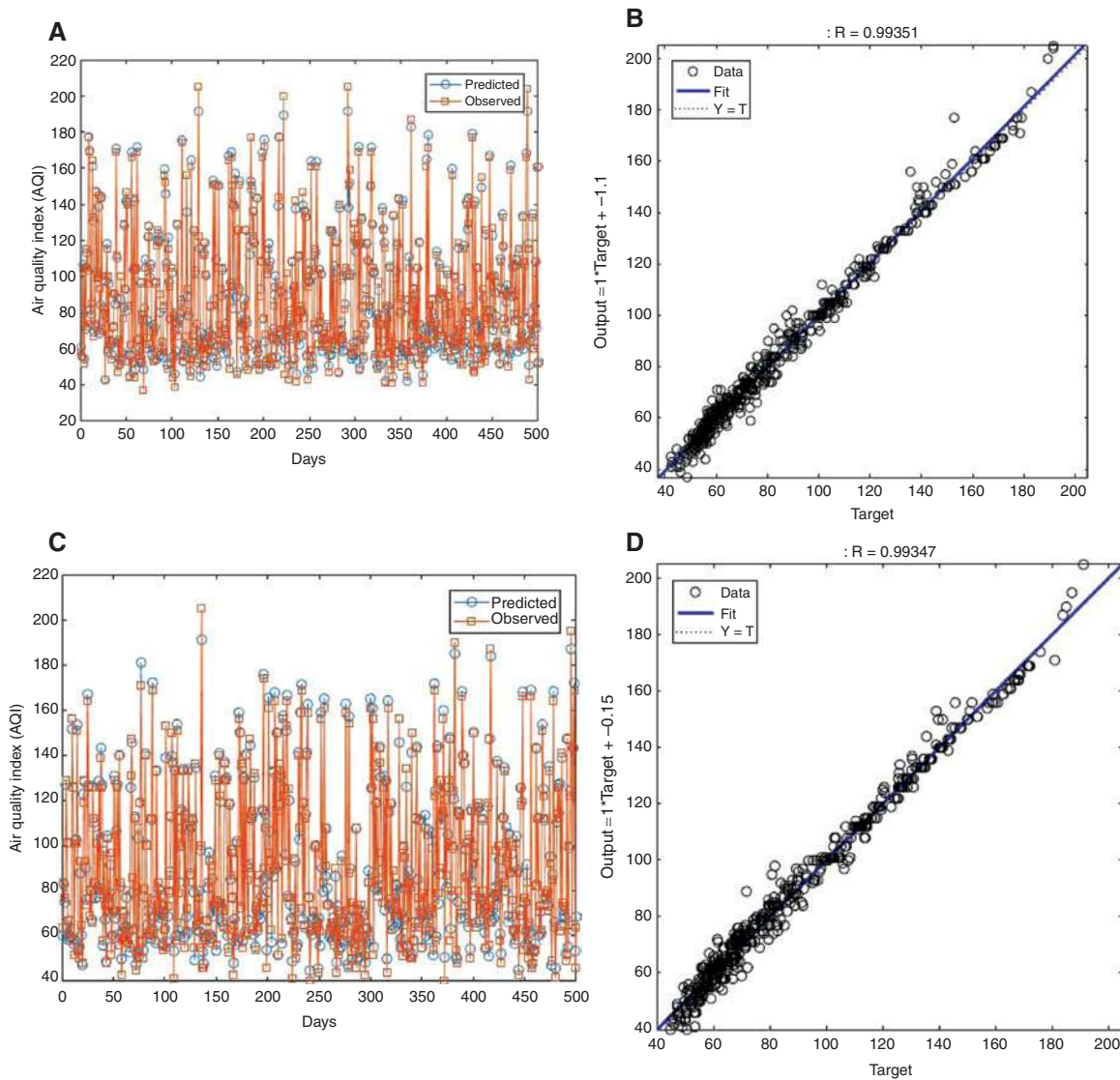
Figure 3: Cascade Forward Ensemble of Neural Network Predictors: (A) Testing Plot for Houston, (B) Regression Plot for Houston, (C) Testing Plot for Los Angeles, and (D) Regression Plot for Los Angeles.



are given as input to cascade forward neural network and SVR for the final forecast of the AQI. The performance of these two combiner algorithms with same base learners is compared.

The testing results of the cascade forward neural network as the combiner algorithm with neural network base learners for forecasting the AQI in Houston and Los Angeles are given in Figure 3. The testing results of SVR as the combiner algorithm with neural network base learners for forecasting the AQI in Houston and Los Angeles are given in Figure 4. The performance measures of both ensembles with neural network predictors as the base learners are given in Table 4.

It can be seen that the cascade forward ensemble outperformed the SVR ensemble of individual neural network predictors. The regression plots represented in Figure 3B and D for Houston and Los Angeles, respectively, indicate that the relationship between the observed and predicted values of the AQI determined by the cascade forward ensemble makes a good fit.

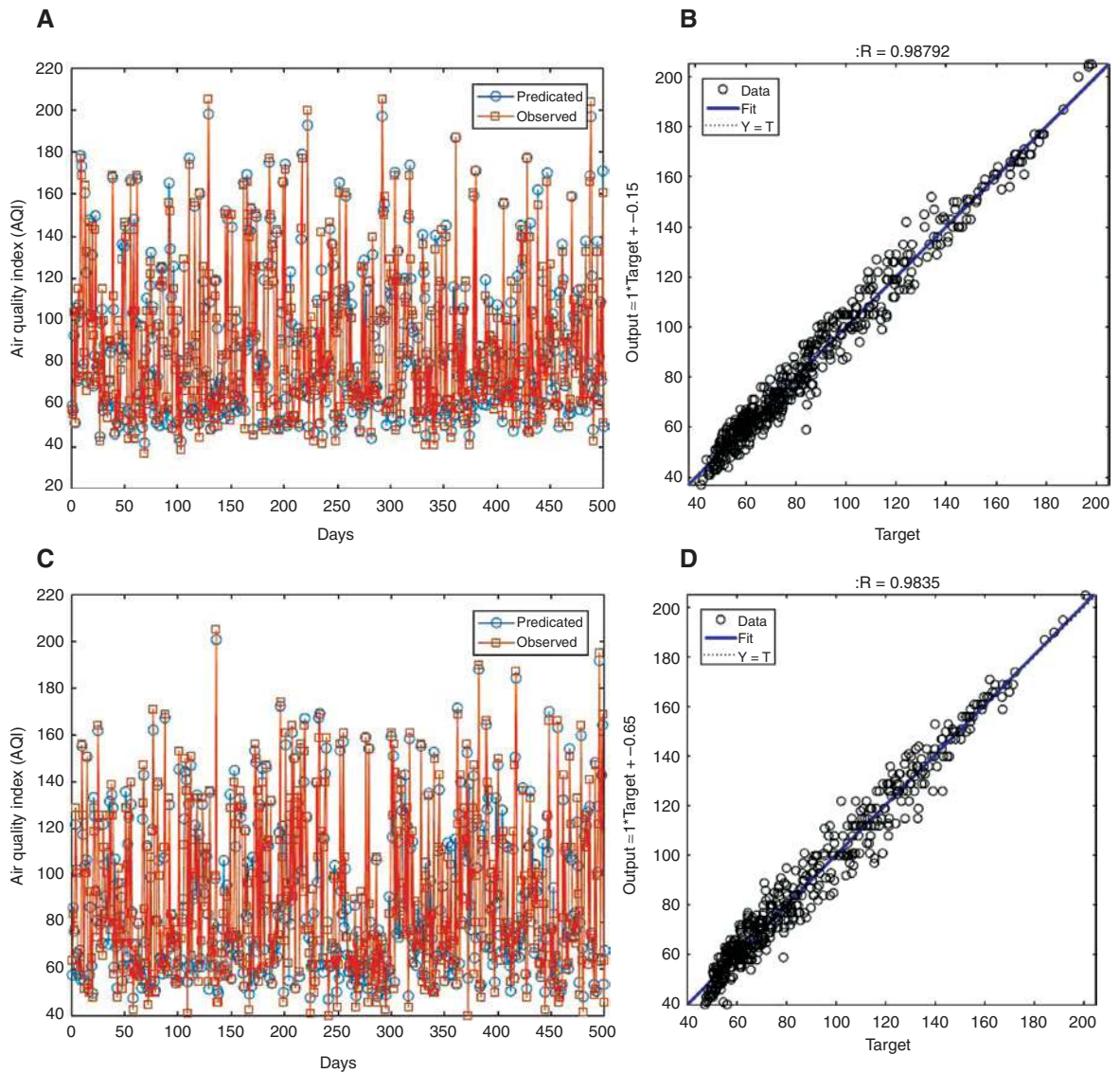


**Figure 4:** SVR Ensemble of Neural Network Predictors: (A) Testing Plot for Houston, (B) Regression Plot for Houston, (C) Testing Plot for Los Angeles, and (D) Regression Plot for Los Angeles.

**Table 4:** Performance of Cascade Forward and SVR Ensembles with Neural Networks as the Base Learners.

Algorithm	MAE		MAPE		R		RMSE		IA	
	H	LA	H	LA	H	LA	H	LA	H	LA
MLP	5.70	8.62	6.76	10.25	0.978	0.947	7.31	11.12	0.989	0.972
Cascade forward	3.04	2.91	3.86	3.69	0.993	0.993	4.06	4.05	0.996	0.996
Elman	9.39	9.04	10.89	10.81	0.939	0.948	12.31	11.23	0.967	0.970
RBF	4.18	4.47	5.38	5.89	0.981	0.981	6.83	6.65	0.990	0.990
NARX	14.35	11.08	17.24	12.71	0.880	0.919	18.06	13.95	0.907	0.951
Ensemble (cascade forward)	2.68	2.70	3.34	3.39	0.994	0.994	3.75	3.70	0.997	0.997
Ensemble (SVR)	2.95	3.01	3.82	3.96	0.993	0.993	4.06	3.96	0.996	0.996

H, Houston; LA, Los Angeles.



**Figure 5:** Cascade Forward Ensemble of Regression Models: (A) Testing Plot for Houston, (B) Regression Plot for Houston, (C) Testing Plot for Los Angeles, and (D) Regression Plot for Los Angeles.

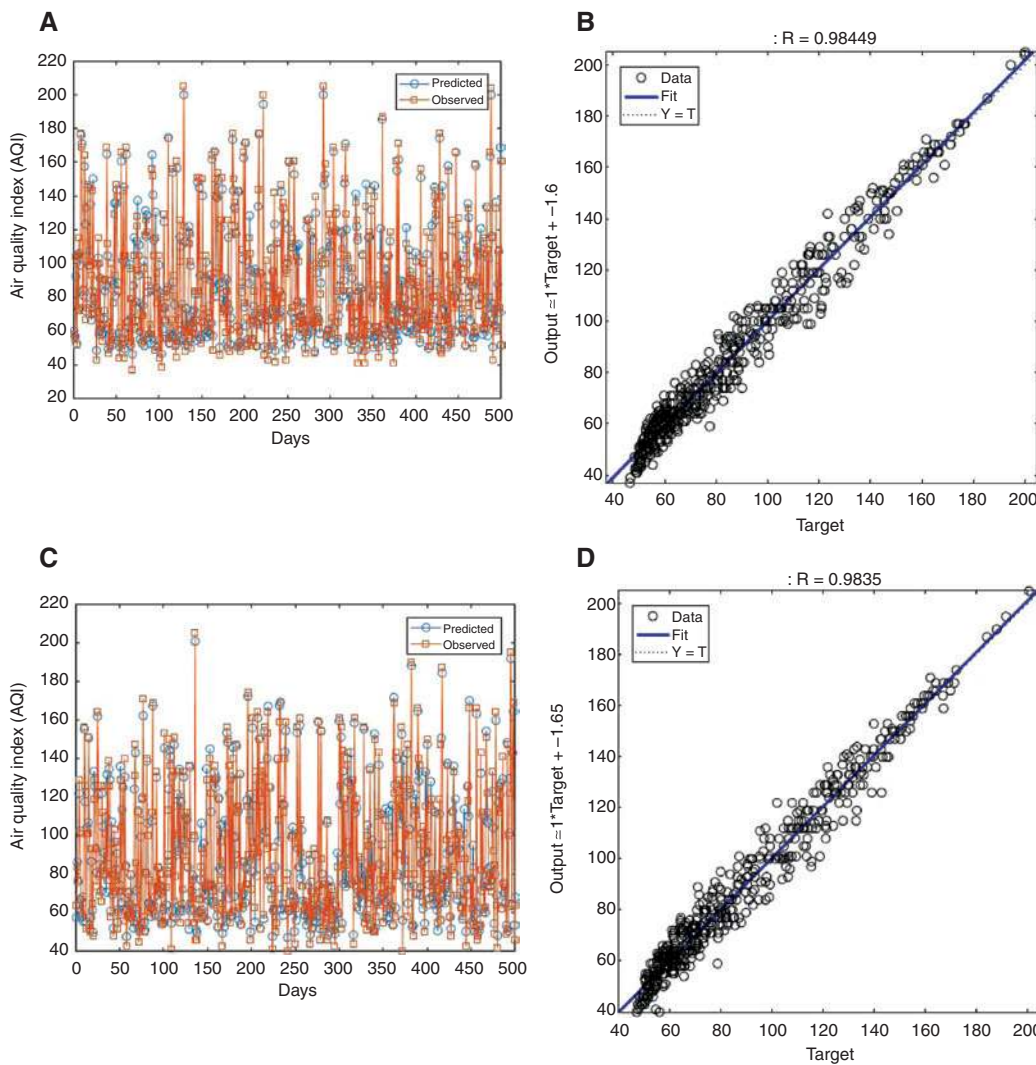


### 3.2 Cascade Forward and SVR Ensembles of Regression Models

In these two ensembles, the base learners are individual regression models: MLR with BGD, SGD, MBDG, and CGD optimization algorithms and SVR. The predictions of these regression models are given as input to the combiner algorithms for the final forecast of the AQI.

The testing results of the cascade forward neural network as the combiner algorithm with regression models as the base learners for forecasting the AQI in Houston and Los Angeles are given in Figure 5. The testing results of SVR as the combiner algorithm with regression models as the base learners for forecasting the AQI in Houston and Los Angeles are given in Figure 6. The performance measures of both ensembles with regression models as the base learners are given in Table 5.

It can be seen that the cascade forward ensemble outperformed the SVR ensemble of individual regression models. Also, most of the neural network predictors exhibited high accuracy compared to MLR models with gradient descent variants. Among the neural network predictors, the cascade forward neural network



**Figure 6:** SVR Ensemble of Regression Models: (A) Testing Plot for Houston, (B) Regression Plot for Houston, (C) Testing Plot for Los Angeles, and (D) Regression Plot for Los Angeles.

**Table 5:** Performance of Cascade Forward and SVR Ensembles with Regression Models as the Base Learners.

Algorithm	MAE		MAPE		R		RMSE		IA	
	H	LA	H	LA	H	LA	H	LA	H	LA
MLR (BGD)	10.46	10.96	12.34	13.06	0.929	0.922	13.06	13.46	0.961	0.957
MLR (SGD)	10.28	10.46	12.11	12.58	0.931	0.929	12.92	12.85	0.962	0.962
MLR (MBGD)	10.21	10.19	12.03	12.37	0.931	0.931	12.90	12.64	0.963	0.963
MLR (CGD)	10.22	10.17	12.05	12.34	0.931	0.931	12.88	12.61	0.963	0.963
SVR	5.91	5.89	7.16	6.99	0.979	0.979	7.25	7.13	0.988	0.988
Ensemble (cascade forward)	4.42	4.22	5.52	5.10	0.987	0.987	5.49	5.45	0.993	0.993
Ensemble (SVR)	5.03	5.00	6.62	6.22	0.984	0.983	6.26	6.30	0.991	0.991

H, Houston; LA, Los Angeles.

has shown high performance, and among regression models, SVR has shown high performance. The cascade forward neural network performed better due to its structure. It has greater ability to map nonlinear input to the output. Also, the neural networks usually outperform regression models such as MLR when nonlinearities are involved. Often, generalization is more in the case of neural network predictors. Both ensembles have shown better accuracy than the individual predictors of the AQI.

## 4 Conclusion

The proposed approach investigates the concentrations of  $\text{NO}_2$ ,  $\text{CO}$ ,  $\text{O}_3$ ,  $\text{PM}_{2.5}$ ,  $\text{SO}_2$ , and  $\text{PM}_{10}$  to build a forecasting model of the AQI. The study has been carried out for Houston and Los Angeles. The accuracy of prediction has been improved by the stacked ensemble of individual predictors. The performance of the implemented ensemble methods for two different sets of base learners has been evaluated and compared. It can be concluded that cascade forward ensemble outperformed SVR ensemble for both sets of base learners. The limitation of the current study is the ability to predict the AQI more accurately from highly nonlinear data. Often, the neural network methods usually converge to locally optimal solutions. The CGD optimization method greatly deals with this problem. The future study can be attributed to the use of deep learning methods such as LSTM and deep belief networks such as restricted Boltzmann machines, as they extract features layer by layer and combine low-level features to form high-level features and they also have the ability to model complex mapping [1]. Fuzzy logic can also be implemented when the behavior of the mathematical model is complicated and the linguistic rules are needed to define the behavior of a system. The scope of the current study also includes the use of ensemble methods such as bagging and boosting. In conclusion, the present study has indicated that the ensemble approach provides promising results for forecasting the AQI compared to individual neural network predictors and regression models.

**Acknowledgments:** The authors would like to thank VIT University (Vellore, India) for supporting the current research work. The authors also acknowledge the US EPA for providing air quality data.

## Bibliography

- [1] Y. Bengion and O. Delalleau, On the expressive power of deep architectures, in: *Proceedings of the 14th International Conference on Discovery Science*, pp. 18–36, Springer-Verlag, Berlin, 2011.
- [2] L. Breiman, Stacked regression, *Mach. Learn.* **24** (1996), 49–64.
- [3] D. S. Broomhead and D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks, Technical report, RSRE, 4148, 1988.
- [4] A. Chaloulakou, M. Saisana and N. Spyrellis, Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, *Sci. Total Environ.* **313** (2003), 1–13.

- [5] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.* **20** (1995), 273–297.
- [6] J. W. Davidson, D. A. Savic and G. A. Walters, Symbolic and numerical regression: experiments and applications, *Inf. Sci.* **150** (2002), 95–117.
- [7] S. Deleawe, J. Kuszniir, B. Lamb and D. J. Cook, Predicting air quality in smart environments, *J. Ambient Intell. Smart Environ.* **2** (2010), 145–154.
- [8] E. G. Dragomir, Air quality index prediction using K-nearest neighbor technique, *Bull. PG Univ. Ploiesti Ser. Math. Inform. Phys.* **LXII** (2010), 103–108.
- [9] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola and V. N. Vapnik, Support vector regression machines, in: M. C. Mozer, M. I. Jordan, T. Petsche, eds., *Advances in Neural Information Processing Systems 9*, NIPS 1996, pp. 155–161, MIT Press, Cambridge, Massachusetts, 1997.
- [10] P. Goyal, A. T. Chan and N. Jaiswal, Statistical models for the prediction of respirable suspended particulate matter in urban cities, *Atmos. Environ.* **40** (2006), 2068–2077.
- [11] N. Gutierrez-Jaime, Population of the City of Los Angeles surpasses 4 million, KTLA 5, Tribune Broadcasting, Retrieved 17 May 2017.
- [12] M. R. Hestenes, R. Magnus and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bureau Stand.* **49** (1952), 409–436.
- [13] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.* **2** (1989), 359–366.
- [14] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, An efficient k-means clustering algorithm: analysis and implementation, *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (2002), 881–892.
- [15] A. Kurt and A. B. Oktay, Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks, *Expert Syst. Appl.* **37** (2010), 7986–7992.
- [16] *New state population report: California grew by over 335,000 residents in 2016*, California Department of Finance. Retrieved 17 May 2017.
- [17] P. Smyth and D. H. Wolpert, Linearly combining density estimators via stacking, *Mach. Learn. J.* **36** (1999), 59–83.
- [18] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios and J. Kukkonen, Evaluation of a multiple regression model for the forecasting of the concentrations of  $\text{NO}_x$  and  $\text{PM}_{10}$  in Athens and Helsinki, *Sci. Total Environ.* **409** (2011), 1559–1571.
- [19] World Health Organization (WHO), *Estimated deaths and DALYS attributable to selected environmental risk factors*, WHO, Geneva, Switzerland, 2015.
- [20] D. Wolpert, Stacked generalization, *Neural Netw.* **5** (1992), 241–259.
- [21] X. Yan, *Linear regression analysis: theory and computing*, pp. 1–2, World Scientific, Singapore, 2009, ISBN 9789812834119.
- [22] G. Zhang, B. Eddy Patuwo and M. Y. Hu, Forecasting with artificial neural networks: the state of the art, *Int. J. Forecast.* **14** (1998), 35–62.