Air Quality Modeling for Sustainable Clean Environment Using ANFIS and Machine Learning Approaches

Osman Taylan
King Abdulaziz University Faculty of Engineering
Abdulaziz Alkabaa (aalkabaa@kau.edu.sa)
King Abdulaziz University Faculty of Engineering https://orcid.org/0000-0001-9016-4241
Mohammed Alamoudi
King Abdulaziz University Faculty of Engineering
Abdulrahman Basahel
King Abdulaziz University Faculty of Engineering
Mohammad Balubaid
King Abdulaziz University Faculty of Engineering
Murad Andejany
University of Jeddah
Hisham Alidrisi
King Abdulaziz University Faculty of Engineering

Research Article

Keywords: Air pollution, Air quality, ANFIS, Big data, Environmental factors, Machine learning

Posted Date: March 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-282971/v1

Air Quality Modeling for Sustainable Clean Environment Using ANFIS and Machine Learning Approaches

Osman Taylan[1], Abdulaziz S. Alkabaa[1,*], Mohammed Alamoudi[1], Abdulrahman Basahel[1], Mohammed Balubaid[1], Murad Andejany[2], Hisham Alidrisi[1]

[1]*Department of Industrial Engineering, Faculty of Engineering, King Abdulaziz University, P.O. Box 80204, Jeddah 21589, Saudi Arabia.*

[2]*Department of Industrial and System Engineering, College of Engineering, University of Jeddah.*

[*]*Corresponding author*

Abstract

Air quality monitoring and assessment are essential issues for sustainable environmental protection. The monitoring process is composed of data collection, evaluation, and decision making. Several important pollution factors, such as SO_2, CO, PM10, O_3, NOx, H_2S, location, and many others, have detrimental effects on air quality. Air quality cannot be precisely recorded and measured due to the total effect of pollutants that usually cannot be collectively prescribed by a numerical value. Therefore, evolution is required to take into account the complex, poorly defined air quality problems in which several naive and noble modeling approaches are used to evaluate and solve. In this study, hybrid data-driven machine learning, and neuro-fuzzy methods are integrated for estimating the air quality in the urban area for public health concerns. 1771 data are collected during three years for each pollution factor, starting from June 1, 2016, till September 30, 2019. The Back-Propagation Multi-Layer Perceptron (BPMLP) algorithm was employed with the steepest descent approach to reduce the mean square error for training the algorithm of the neuro-fuzzy model. Levenberg-Marquardt (LM) approach was also employed as an optimization method with Artificial neural networks (ANNs) for solving nonlinear least-squares problems in this study. These approaches were evaluated by fuzzy quality charts and compared statistically with the US-EPA air quality standards. Due to the effectiveness and robustness of soft computing intelligent models, the public's early warning will be possible for avoiding the harmful effects of pollution inside the urban areas, which may reduce respiratory and cardiovascular mortalities. Consequently, the stability of air quality models was correlated with the absolute air quality index. The findings showed remarkable performance of ANFIS and ANN-based Air Quality models for High dimensional data assessment.

*Keywords***:** Air pollution; Air quality; ANFIS; Big data; Environmental factors; Machine learning.

1. Introduction

One of the most critical factors that significantly affect climate change and human health is air pollution. Many countries have been using different systems for monitoring air pollution. Thus, this area of research is of interest and very active. Several naive and noble modeling approaches have been presented in the literature, such that hybrid approaches (Zhu et al., 2018), and linear unbiased estimator (Sozzi et al. 2017), an autoregressive integrated moving average (ARIMA) (Reikard, 2019), bias adjustment (Silibello et al., 2015), principal component regression approach are naive methods. Nevertheless, non-parametric regression (Donnelly et al., 2015), Artificial Intelligence (AI) techniques, machine learning (Rybarczyk et al., 2018), neuro-fuzzy approaches are noble air quality modeling and control systems. Similarly, simulation and data mining are well-known modeling tools and techniques for predicting and assessing air quality. In this context, Aggarwal et al. (2019) and Bai et al. (2017) have concentrated on the models used to predict the abnormality exploration in air quality. Deep learning applications (as a subset of machine learning) have recently shown considerable potential for investigating further aspects of the ecological dimensions (Christin et al., 2019; Fairbrass et al., 2019; and Torney et al., 2019). A recent study by Sayeed et al. (2019) proposed an artificial intelligence (AI) model using deep convolutional ANNs to predict 24 hours ozone concentration in Texas for comparing the results of different periods in the year 2017. Munawar et al. (2017) presented a case study of Lahore city of Pakistan for the prediction of Air Quality Index (AQI) using a hybrid approach of neuro-fuzzy inference systems. Rahman et al. (2020) investigated the soft computing applications of air quality modeling by reviewing and discussing the neuro-fuzzy systems, fuzzy logic, deep learning, conventional and evolutionary ANNs, and many hybrid models. Hvidtfeldt et al. (2019), Ansari and Ehrampoush (2019), and Liu et al. (2019) expressed the exposure to pollutants causes different diseases such as respiratory diseases, asthma, type 2 diabetes, cancer, and allergies. Alimissis et al. (2018), Cabaneros et al. 2019), and Taylan (2017) searched the air quality models playing crucial roles to evaluate the air quality problems in the atmosphere. These models can show the health conditions in the cities using the domain knowledge using reliable and noble forecasting approaches. The advantages of these models are that they can provide early warning in case they are effectively utilized and can reduce the numbers of manual measurements of data acquisition substantially. As a modeling approach, ANNs provide effective, flexible, less assumption dependent outcomes. They have adaptive properties and can be integrated with other modeling approaches to assess and control environmental systems. The integration of ANNs and fuzzy logic models called neuro-fuzzy modeling approaches have obtained extensive attention in air quality modeling due to their adaptiveness and well-generalized performance. The different

potentials of ANNs have been employed for modeling the various air pollutants, including NOx and SOx (Radojevi'c et al., 2019), COx, O3 (Pawlak et al., 2019), PM10 (Biancofiore et al., 2017), and PM2.5 in different places all over the world. In this context, Grivas and Chaloulakou (2006) used evolutionary computational algorithms such as ANNs in air quality modeling; similarly, they used genetic-algorithm-tuned ANN hybrid models for the hourly PM10 concentrations in Greece.

Similarly, as an evolutionary approach, fuzzy modeling is used to deal with the vagueness and uncertainties of real-world problems using fuzzy *'If-Then'* rules. A rule set is designed to control the possible relations between the input and output factors by a fuzzification process. Fuzzy modeling is a robust tool to solve complex engineering problems that are difficult to solve by traditional algebraic models. These modeling approaches encapsulate the vagueness of linguistic parameters and terms of qualitative factors. Jorquera et al. (1998) demonstrated the usefulness of fuzzy logic modeling in predicting the maximum daily O3 concentration levels. The adapted neuro-fuzzy and fuzzy logic approaches have been used to model concentrations of O3 and PM10. Ghoneim et al. (2017) and Zhou et al. (2019) employed deep learning and deep multi-output long short-term memory ANNs models for determining the air pollutants' concentration. Habaneros et al. 2019) (Rybarczyk et al., 2018) claimed that only a few review articles are available to discuss the soft computing techniques in air quality modeling (Ayturan et al., 2018, Zhou and Xie, 2020; Iskandaryan et al. 2020) where it was found that these ANNs or/and deep learning techniques are mostly limited applications. The articles covering the whole spectrum of the available soft computing techniques can rarely be found.

The state of air pollution is frequently expressed by Air Quality Index (AQI). AQI is extensively used for air quality assessment and management (Sowlat et al., 2011). USA Environmental Protection Agency and local authorities use AQI to provide air quality information of a location and its impact on health (EPA, 2005). High AQI values mean increased pollution and living things started to begin exposition of health problems (US EPA, 1999). Sulphur Dioxide (SO2 μg/m3), Carbon Monoxide (CO mg/m3), Particular matters (PM10 μg/m3), Ozone (O3, μg/m3), and Nitrogen oxide NO μg/m3), Hydrogen sulfur H2S (μg/m3) are considered pollutants in the urban area. The AQI categories and their standard quality intervals are given in Table 1. These categories of AQI have been identified by fuzzy linguistic terms and their numerical intervals for air quality assessment.

Table 1. Air Quality Index (AQI) categories (EPA-US-2005)

AQI categories	Quality levels of health concern
0-50	Good
51-100	Moderate
101-150	Unhealthy for sensitive groups
151-200	Unhealthy
201-300	Very unhealthy
>301	Hazardous

In this study, initially, statistical inferencing approaches were used to examine the underlying relationship between the pollutants and their impacts on the air quality index. Eq. 1 is a way to present the relationship between an air pollutant concentration and AQI. The pollutant concentration in this equation was defined as a ratio of the relevant standard.

$$\text{Air quality index} = \frac{C_i}{S_i} 500 \quad (1)$$

Where c_i and s_i show the pollutant concentration and standard pollutant level, respectively. In recent years, several research types were carried out to develop air quality prediction models for launching ambient air quality standards. Numerous guidelines have been presented to set the level of air quality bounds on the emissions of pollutants (US EPA, 1999). On the other hand, determining and developing AQ limits using big data is a very recent work. Attention was mainly given to soft computing techniques to obtain and evaluate the big data (Kaur et al., 2017) regarding the air quality models. Due to the size and complexity of big data in air quality systems, the essential for soft computing approaches have extensively increased, particularly with the growing interests in the systems of early warning alerts and preventive actions for pollutants' when high concentrations of pollutants are observed (Taylan, 2017). Recently, several attempts have been conducted to investigate air quality using machine-learning and neuro-fuzzy (ANFIS) approaches and big data analytics (Masmoudi et al., 2020; Sharma et al., 2020; Aggarwal et al., 2019; Macing et al., 2019; Sayeed et al., 2019; Bai et al., 2017; Pan et al., 2017; Wang et al., 2017; Prasad et al., 2016).

The characteristics of modeling approaches require different types of data sets. For instance, ANNs and fuzzy systems are bidirectional and need numerical and linguistic data, which are broadly discussed (Ishibuchi et al., 1999). On the other hand, fuzzy systems can organize, handle, and use vague, imprecise, and uncertain information, construct balance among different and inconsistent observations, and use subjective and qualitative information to model complex problems (El Raey, 2006). As seen in Table 1, linguistic terms are employed for air quality assessment together with

numerical values. The numerical data shows the upper and lower limit of parameters that the observations have taken place. Taylan et al. (2009) used numerical data to train machine learning approaches and develop adaptive fuzzy models using symbolic qualitative and numerical data. Neuro-fuzzy systems integrate neural networks and fuzzy systems for developing models that have learning capabilities obtained through training processes. The goal of hybrid integration with big data is to form a more intelligent system to predict and control the air quality. However, applying a hybrid neuro-fuzzy system is very rare in air quality prediction and control systems. These hybrid approaches can predict the air quality, evaluate the findings, and provide online information. In case of unhealthy or hazardous conditions, local authorities can take immediate public actions more intelligently. In this study, the modeling method considers six major air pollutants as input parameters; SO2, CO, PM10, O3, NO, H2S, and the output parameter is the AQI. 1771 data were obtained for each parameter, 1065 data (60%) were used for training, 353 data (20%) were used for testing, and the remaining 353 data (20%) were intended for the validation of the model.

These steps of the modeling approach are presented in detail in section 2. The article is organized as follows: Section 2.1 describes the significant air pollution sources and their impacts on the air quality index. Section 2.2 shows the application of ANFIS in air quality modeling. The details of ANFIS modeling were presented in section 2.3. Section 3 explains the machine learning approach for air quality estimation. The results and discussions are given in section 4. Finally, the research ends with conclusions and references.

2. Materials and methods

2. 1. Major sources of air pollution and their impacts on air quality

Several factors affect air pollution, such as dust storms, particulate matter, greenhouse gases, other gas emissions, urban growth, and transportation. The impacts of sulfur dioxide, nitrogen dioxide, and ozone cause declines in crop yields and affect human health (El Raey, 2006). Alternatively, ozone is caused by complex chemical reactions in the atmosphere (Al-Alawi et al., 2008). The highest level of pollution occurs where pollutant concentrations are the greatest. The level of pollutions allowed is given in Table 2, where air quality standards in Saudi Arabia, Gulf countries, and US-EPA are presented.

Table 2. Air quality standards in Saudi Arabia, Gulf countries and US-EPA

	Air quality standards

Air Pollutant	KSA	Gulf Countries	US-EPA Standards
Sulfur dioxide (SO_2)	730 μg/m^3 (1 h) 365 μg/m^3 (24 h) 85 μg/m^3 (1 year)	441 μg/m^3 (1 h) 217 μg/m^3 (24 h) 65 μg/m^3 (1 year)	80 μg/m^3 (annual arithmetic mean) 365 μg/m^3 (24h average)
Nitrogen oxides NO_2	660 μg/m^3 (1 h) 100 μg/m^3 (1 year)	660 μg/m^3 (1 h) 100 μg/m^3 (1 year)	100 μg/m^3 (annual arithmetic mean)
Ozone (O_3)	295 μg/m^3 (1 h)	235 μg/m^3 (1 h) 157 μg/m^3 (8 h)	235 μg/m^3 (1h average) 157 μg/m^3 (8h average)
Carbon monoxide (CO)	40000 μg/m^3 (1 h) 10000 μg/m^3 (8 h)	40000 μg/m^3 (1 h) 10000 μg/m^3 (8 h)	10 μg/m^3 (8h average) 40 μg/m^3 (1h average)
Hydrogen sulphide (H_2S)	200 μg/m^3 (1 h) 40 μg/m^3 (24 h)	200 μg/m^3 (1 h) 40 μg/m^3 (24 h)	200 μg/m^3 (1 h) 40 μg/m^3 (24 h)
Particulate matters (PM10)	340 μg/m^3 (24 h) 80 μg/m^3 (1 year)	340 μg/m^3 (24 h) 80 μg/m^3 (1 year)	50 μg/m^3 (annual arithmetic mean) 150/m^3 (24h- average)

Ozone and sulfur dioxide are considered the leading causes of the low yield of crops because of the acidification of soils, lakes, and streams. When the soils are acidified, acidity and toxic aluminum move from catchments into lakes and sea, making them highly polluted. The nitrogen disordering can acidify the soil, fertilize sensitive natural plant communities, and cause irregularity that can affect imbalance ecosystems. Fig. 1 (a), (b), (c), and (d) illustrate the effects of pollutants: sulfur dioxide, particulate matter, ozone, and hydrogen sulfur on air quality. The figure shows that an increase in sulfur dioxide, particularly nitrogen oxide, raises the AQI, which means a high level of pollution and reduced air quality.

On the other hand, the effect of carbon monoxide is more complicated; this gas is a toxic air pollutant, mainly produced largely from vehicle emissions which have health effects including weakness, vomiting, headaches, nausea, clouding of consciousness, coma, and unfortunately at high concentrations and long enough exposure cause death. It also raises the AQI and reduces the air quality. However, this study aims to find out the cumulative effect of pollutants on air quality.

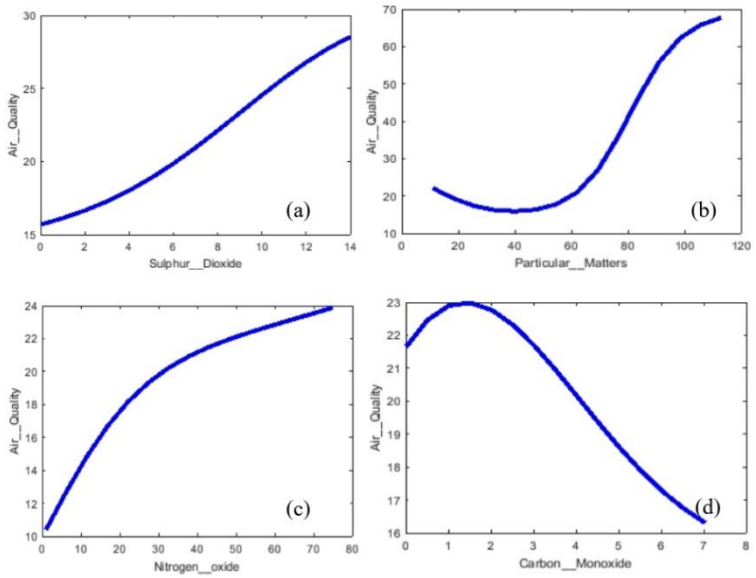

Fig.1. The impacts of pollutants on the air quality

Air pollutants encounter the human body mainly via the respiratory system. Ozone, NO, and SO2, delicate particulate matters, and dust can affect the mucous membranes' inflammation. These redden the eyes, inflame the pharynx and throat, red lung functions, and weakness the immune system, which eventually cause respiratory diseases. Several symptoms may occur, such as headaches, giddiness, nausea, and pounding of the heart as the signs of extreme exposures. The US-EPA (1999) standards were considered for the conversion of pollutants' data into the indexes. As shown in Table 1, when AQI is between zero and 50, the level of health concern is good for society. Conversely, higher AQI means high-level pollution, which is risky for public health.

2. 2. Application of ANFIS for air quality modeling

An ANFIS model designed with suitable input-output parameters can depict a human expert's behaviours to control the air quality between the predefined parameters. The model can use environmental data, produce suitable outcomes of AQI and inform authorities. An adaptive network is connected by links, where each node executes a function on incoming signals from sensory information of pollutants to produce output and specifies the direction of signal flow between the nodes (Jang et al., 1997). In a typical network, nodes present mathematic functions modifiable by specified parameters. These parameters can impact the performance of the network and its functions. However, in this work, the mathematical functions are replaced with fuzzy rules. As shown in Fig.2, membership functions take the place of mathematical equations and carry out their duties, making this approach unique and noble for air quality modeling. The complete fuzzy rules set given below,

is the backbone of the expert system. Fig. 2 shows the architecture of the ANFIS model for the prediction of the air quality index. An ANFIS model consisting the fuzzy if-then rules (Rs) is a fundamental tool for assessing the air quality. The input parameters are X_i = {*x_1: Sulphur dioxide* (SO_2)*, x_2: Carbon monoxide (CO), x_3: Hydrogen sulfide (H_2S), x_4: Ozone (O_3), x_5: Nitrogen oxide* (NO_x)*, and x_6: Particulate matters (PM_{10})*}. The output parameter is the air quality Index (y_i; AQI). The rules are the backbone of ANFIS model, their MFs (μs) are Gaussian functions depicting the fuzzy linguistic terms (φs) and are presented in the rule set given below.

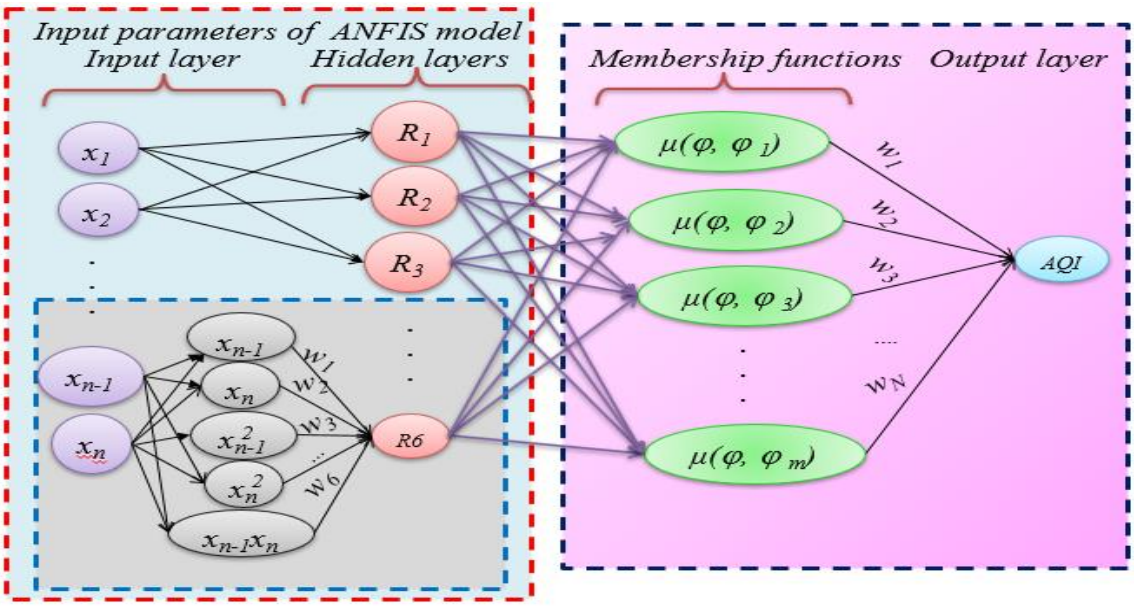

Fig. 2. The ANFIS model architecture for air quality prediction

It is essential to mention that there are often uncontrollable and unavoidable causes of variations in air quality. Identifying variations require dealing with air quality characteristics using linguistic terms. Collecting numerical data about the air pollutants is essential, but this will not be as meaningful as linguistic terms used to identify the air quality parameters. Because crisp numbers cannot identify some parameters, fuzzy linguistic terms might be more suitable to deal with these parameters. For instance, air quality is a linguistic variable whose values might be linguistic terms as good, healthy, unhealthy, very unhealthy, hazardous, etc. Due to the imprecision and vagueness in these quality measures, a trend was initiated to integrate the randomness and fuzziness for assessing environmental quality problems. In fig. 3 (a) and 3 (c), the air quality index is plotted in three-dimensional (3D) graphs versus Carbon monoxide and sulphur dioxide. Similarly, it was plotted against ozone and nitrogen oxide for Jeddah, respectively. Fig. 3 (b) and 3(d) show that nonlinear relation appears clearly between the input parameters and air quality index. The 3D plots are very obliging for observing the full view of the air quality index's output surface based on the

whole span of the input parameters. The 2D and 3D plots of air quality index and regressors such as ozone, sulphur dioxide, carbon monoxide, and nitrogen oxide showed that the system was nonlinear and recommended the evolution of an intelligent approach to predict and control the air quality in a city.

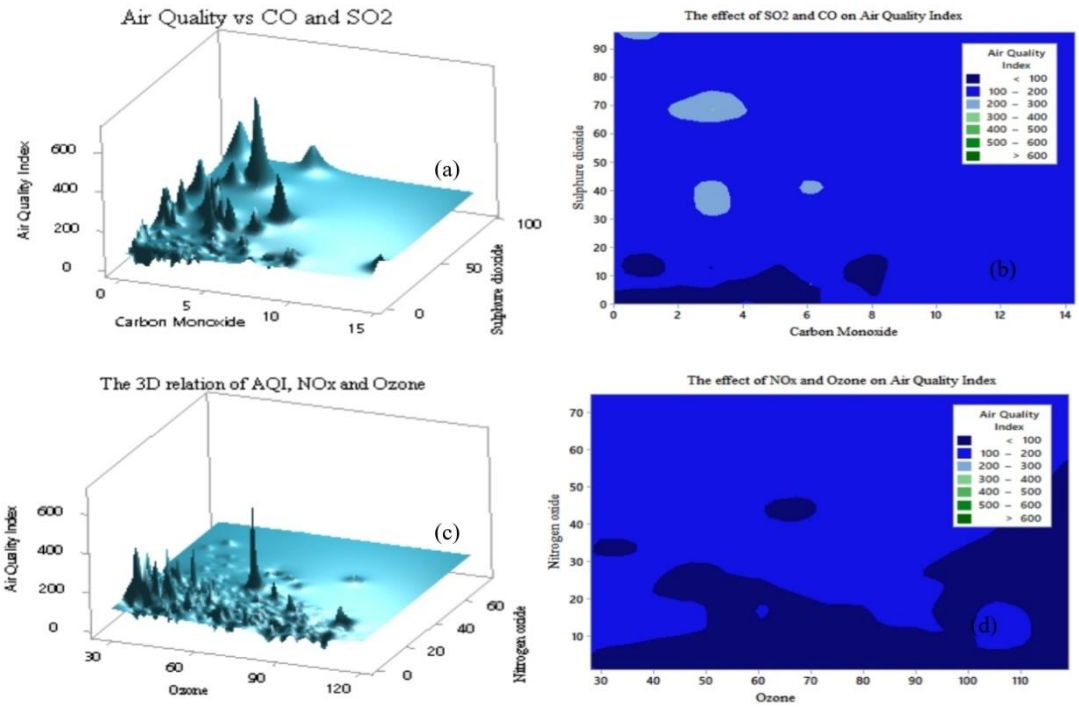

Fig. 3. The impacts of pollutants on the air quality index

The analysis of 3D surfaces shows that many local maximum and minimum points appear in the responses of the given parameters. Therefore, this reveals that the rise (or maximum points) in the pollutant concentration will increase the AQI and cause many negative effects. On the other hand, the local and global minimum points show where the AQI is low, and air quality is good and healthy. Hence, a highly nonlinear relation appears between the pollutants and air quality index.

2. 3. ANFIS based reasoning for air quality prediction.

The ANFIS model was established from six rules and the linguistic statements arranged for air quality modeling and prediction. Fuzzy rules are used to map input parameters to the output. A fuzzy rule is constituted from two parts: the assertion and the conclusion parts, including linguistic variables and their term sets. A clustering analysis was carried out, and the optimal number of clusters was found six with 99.9480% similarity level and 0.00104 distance level between the

clusters. Therefore, rules number were considered equal to the number of clusters; each rule represents the characteristic of data in the cluster for identifying the AQI. Due to the nonlinearity (see in Fig.1 and 3), Gaussian membership functions (MFs) (see in Fig. 4 and 5) were employed for the fuzzy input sets and delta functions for the output spaces. In this study, the centre average defuzzification and product premise approach were employed for obtaining the outcomes of AQI, as given in Eq.2.

$$f\left(\frac{x}{\theta}\right) = \frac{\sum_{i=1}^{R} b_i \prod_{j=1}^{n} exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]}{\sum_{i=1}^{R} \prod_{j=1}^{n} exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]} = \frac{\sum_{i=1}^{R} b_i . \mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)} \quad (2)$$

$$\mu_i(x) = \prod_{j=1}^{n} exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]$$

R represents the number of rules in the rule base, and n denotes the number of inputs per data tuple. θ is represented in a vector form that contains the MF parameters for the rule base c_i, σ_i, and b_i. In the rule base, the Gaussian MFs are used for the rules' premises part, and the delta function is for the conclusion part. The coefficient b_i represents the point in the output space at which the output MF for the ith rule is a delta function and denotes the point in the j^{th} input universe of discourse, where the MF for the ith rule achieves a maximum. It is essential to mention that the relative width; of the j^{th} input MF for the i^{th} rule is always larger than zero. Fuzzy reasoning is the crucial factor in the modeling of fuzzy set theory. For the prediction of air quality, the input membership functions, fact base, the ruleset, and the inference engine are presented in Fig. 4. These fuzzy rules and the reasoning process and defuzzification are considered the pillar of the fuzzy inference system to obtain the outcomes of a model. Fig. 4 shows the fuzzy reasoning procedure of the Sugeno fuzzy model for predicting the air quality in Jeddah.

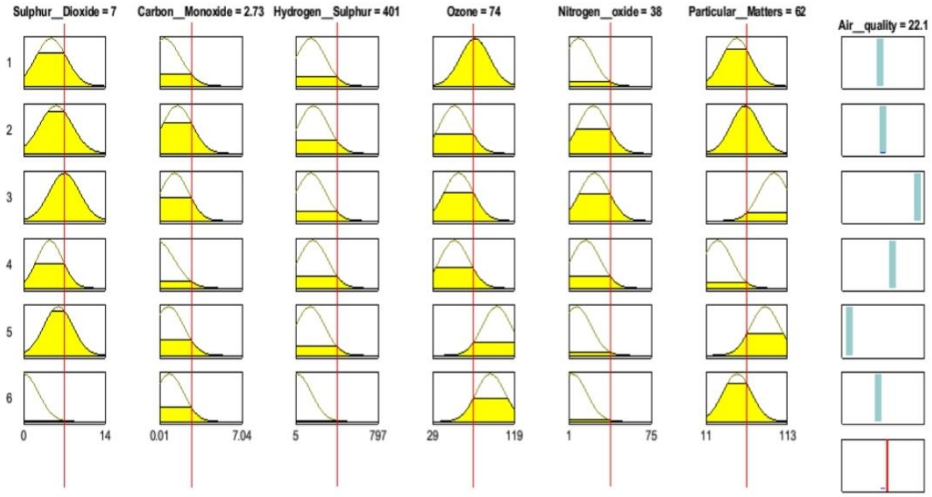

Fig. 4. Fuzzy reasoning procedure for predicting air quality

If air pollution is considered a space-defining by fuzzy set *U, X_is* is the fuzzy input parameter in this space and Y_i is the fuzzy output parameter. Hence, The input parameters of this work are; $X_i = \{x_1$: *Sulphur dioxide* (SO_2)*, x_2: Carbon monoxide (CO), x_3: Hydrogen sulfide (H_2S), x_4: Ozone (O_3), x_5: Nitrogen oxide* (NO_x)*, and x_6: Particulate matters (PM_{10})*}. The output parameter is the air quality Index (y_i; AQI). The fuzzy linguistic term set; $V = V_1 x\, V_2 x \ldots x V_m$, employed for this study is; *V= {good, moderate, unhealthy, very unhealthy and hazardous}*. A fuzzy model is structured by the collection of fuzzy *If-Then* rules. The upper and lower limits of all input parameters and output are presented in Fig. 4. This figure also depicts the fuzzy reasoning procedure. For instance, the upper and lower bounds of *Sulphur dioxide* (SO_2) are between 0-14 µg/m^3, ozone's (O_3) is between 29-119 µg/m^3, and particulate *matters (PM_{10})* is between 11-113 µg/m^3, and so on. The membership functions $\mu_i(x)$; $i = 1, 2, \ldots, n$, are always parametric functions used in the fuzzy model. Fig. 5 (a), (b), (c), and (d) depict the MFs and their term sets of sulphur dioxide, ozone, particulate matters, and nitrogen oxide in Jeddah, respectively.

Rule 1: *IF* (SO_2) *is low and (CO) is low and (H_2S) is low and (O_3) is low and* (NO) *is low and (PM_{10}) is low THEN The air quality is good.*

Rule 2: *IF* (SO_2) *is low and (CO) is normal and (H_2S) is normal and (O_3)is low and* (NO) *is normal and (PM_{10}) is normal THEN The air quality is good.*

Rule 3: *IF* (SO_2) *is high and (CO) is normal and (H_2S) is normal and (O_3)is low and* (NO) *is normal and (PM_{10}) is very high THEN The air quality is normal.*

Rule 4: *IF* (SO_2) *is low and (CO) is high and* (H_2S) *is high and* (O_3)*is very low and* (NO) *is high and* (PM_{10}) *is very low THEN The air quality is unhealthy.*

Rule 5: *IF* (SO_2) *is normal and (CO) is high and* (H_2S) *is high and* (O_3)*is very high and* (NO) *is high and* (PM_{10}) *is high THEN The air quality is unhealthy.*

Rule 6: *IF* (SO_2) *is very low and (CO) is very high and* (H_2S) *is very high and* (O_3)*is high and* (NO) *is very high and* (PM_{10}) *is low THEN air quality is hazardous.*

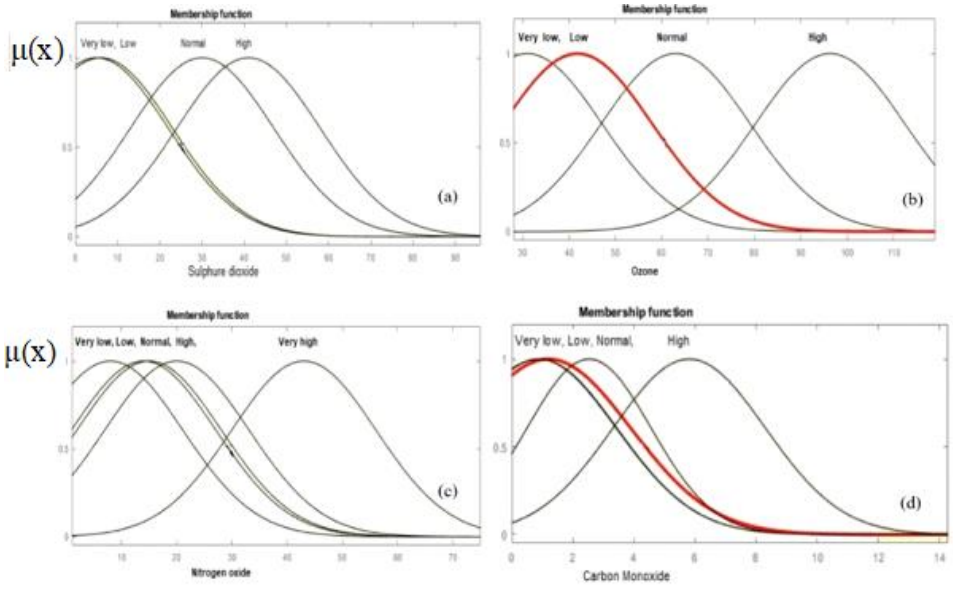

Fig. 5. Membership functions and their terms set for air quality parameters.

Appropriate fuzzy separation of input and output data spaces and a correct choice of MFs are the essential part obtaining a useful ANFIS model for AQI. The MFs and the fuzzy term sets of all variables are determined based on the domain knowledge of the system considered. The Gaussian MFs are identified by two parameters (c, σ), where c denotes the MFs' center, and σ represents the MFs width. Fig. 5 (c) shows the Gaussian MFs for 'Nitrogen oxide' and fuzzy term set 'normal' representing the MFs. Some other fuzzy variables and their MFs are presented in Fig 5. For example, the MF of 'Nitrogen oxide' for the fuzzy term 'normal' is mathematically presented as given in Eq. 3.

$$gaussian(x,c,\sigma) = e^{-1/2\left(\frac{x-c}{\sigma}\right)^2} \tag{3}$$

$$\mu_A(Nitrogenoxide) = \mu_{Normal} = \begin{cases} 0 & for \quad x\langle 5 \, and \, x\rangle 75 \\ e^{-1/2\left(\frac{x-45}{70}\right)^2} & for \quad 5 \le x \le 75 \end{cases}$$

Big data set was used for the training, testing, and validation of the ANFIS model developed which can cover the nonlinear functional dependency between the input and output parameters. The root-mean-square error (RMSE) approach was employed for the error determination, in which 'o_i' and

'p_i' are the observed and predicted values of error, respectively, for the AQI. Eq. 4 gives the mean square error of the ANFIS model developed for this study.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2} \quad (4)$$

Fig. 6 shows the relative error of training and testing data determined for the ANFIS model developed.

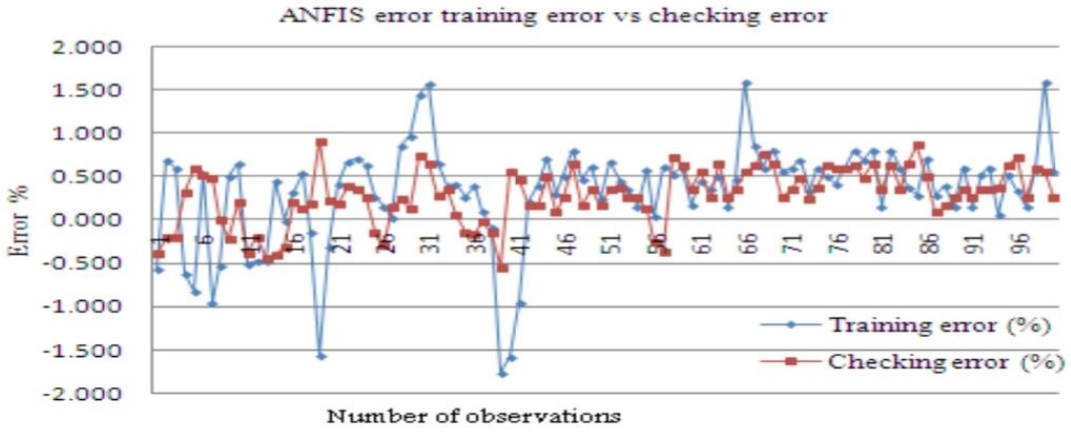

Fig. 6. The training and checking error were determined for the ANFIS model.

As seen in Fig. 6, the relative errors are tolerable and the model checking performance is good. On the other hand, the average training error was found 4.42, and the RMSE for the training data set was calculated as 5.64. Essentially, both the RMSEs are very small for the training and testing of the ANFIS model. Therefore, the developed ANFIS identifies the essential components of the underlying dynamics. In the backpropagation learning algorithm, η and μ are used for 'speeding up' or 'slowing down' the error convergence established in the range of '0' and '1'. The performance of the ANFIS model is presented in Table 3. In case these errors exceed the statistical standards (the '*d*' value), the network is retrained with the increased number of epochs with a repeating process. The magnitudes of 'd' are not the measure of correlation but rather the error's predicted model outcomes. It takes values between 0 and 1; the perfect agreement between the observed and predicted values is when '*d*' is '1', however '0' means absolute disagreement. The value of 'd' can be calculated as given in Eq. 5 follows:

$$d = 1 - \frac{\sum_{i=1}^{N}(p_i - o_i)^2}{\sum_{i=1}^{N}[|p_i - \bar{o}| + |o_i - \bar{o}|]^2} \quad (5)$$

Where $\bar{o}$ represents the observed data average, and $'p'$ is the predicted data.

Table 3. The parameters for determining the strength of the ANFIS model.

Epoch	Number of fuzzy rules of ANFIS model	Statistics *'d'* RMSE		Mean square error after model stabilization (%)
3000	3	0. 527	12.634	0.531
	5	0.351	6.768	0.469
	6	0.285	1.528	0.244
	11	0.491	7.936	0.328
	15	0.648	8.604	0.375
	17	0.692	10.486	0.479
	20	0.592	11.943	0.527
	21	0.727	15.631	0.684
	25	0.731	17.859	0.725

3. Machine Learning Approach for air quality estimation

ANNs are computing systems capable of deep learning and are made up of several highly interconnected elements for information processing. In this work, a back-propagation multilayer perceptron (BPMLP) algorithm was employed for estimating the air quality (y_i) level in Jeddah city. The BPMLP algorithm can perform certain nonlinear mapping that can be described by the terms for a given set of input parameters such as *Sulphur dioxide* (SO_2)*, Carbon monoxide (CO), Hydrogen sulfide* (H_2S)*, Ozone* (O_3)*, Nitrogen oxide* (NO_x)*, and Particulate matters* (PM_{10}). The big data set was divided into suitable parttions for training process, after fifteen iterations appearing in Fig.7, considering the weight distribution and its allocation, the minimum error was obtained using the mean square error approach.

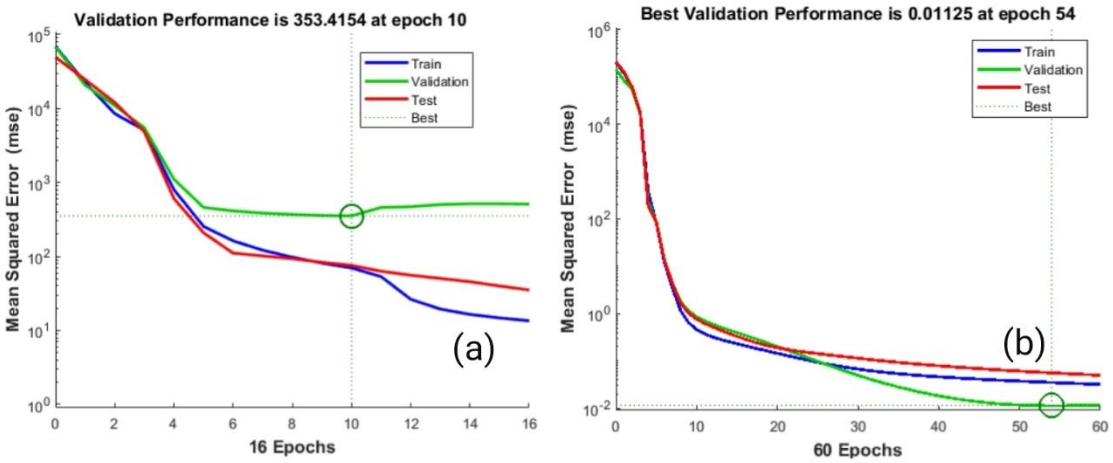

Figure 7. The weight distribution and allocation of training, testing, and validation for

obtaining the minimum error.

The problem of nonlinear relation minimization was solved by the Levenberg-Marquardt (LM) algorithm. The algorithm of steepest descent is known as the error Backpropagation (EBP) Algorithm and is considered one of the most crucial parts in the implementation of training the machine learning algorithm. However, this algorithm's disadvantage is the slow convergence, which can be significantly enhanced by applying the Gauss-Newton algorithm. In evaluating the error surface's curvature, it is customary to use the second-order derivatives of the error function. The Gauss-Newton algorithm can be employed for obtaining the suitable step sizes for each direction and rapidly reach the convergence. As seen in fig. 7 (a), the error function seems to have a quadratic surface. In the initial iteration, the learning is weak (see Fig. 7(a), and the error rate is high. After some iterations (see fig.7(b)), the algorithm could converge quickly and directly. Hence, the learning level is high, and the error rate is low. LM algorithm integrates two minimization methods: the Steepest Descent method and the Gauss-Newton algorithm, for fitting the error curve. However, combining these two algorithms reduces the variance by simultaneously updating the parameters in the Steepest Descent direction.

3.3. Levenberg –Marquardt (LM) Algorithm

As the Hessian matrix $\boldsymbol{J^T J}$ is invertible, the LM algorithm presents another approximation to the Hessian matrix presented in Eq. (6).

$$H \approx J^T J + \delta I \tag{6}$$

Where δ is an always positive combination coefficient, and $\boldsymbol{I}$ is the identity matrix in Eq. (6), in which the elements of the Hessian matrix is greater than zero and is always invertible. The Hessian matrix appearing in Eq. (6) is updated is presented as in Eq. (7).

$$w_{k+1} = w_{\mathrm{k}} - \left(J_k^T J_k + \delta I \right)^{-1} J_{\mathrm{k}} e_k \tag{7}$$

As the LM algorithm integrates the steepest descent and the Gauss-Newton algorithms, it switches between the two algorithms during the training process and gains both advantages. Selecting a very small (nearly zero) combination coefficient δ, Eq. (7) is updated and the Gauss-Newton algorithm is employed to implement the LM algorithm for the training of data obtained from the set of input parameters including *x_1: Sulphur dioxide* (SO_2)*, x_2: Carbon monoxide (CO), x_3: Hydrogen sulfide (H_2S), x_4: Ozone (O_3), x_5: Nitrogen oxide (*NO_x*), and x_6: Particulate matters (PM_{10})*, and the output parameters if *'AQI.'*

As seen in Fig 8, with ANNs, two problems have to be solved; the calculation of the Jacobian

matrix, and organization of the training process. Considering the neuron *j* with *ni* inputs in the first layer, as seen in fig. 8, all its independent parameters are connected to the network's input layer. Eq. 8 was employed to calcuate the air quality index given in the neuron *j* as the output of the ANN.

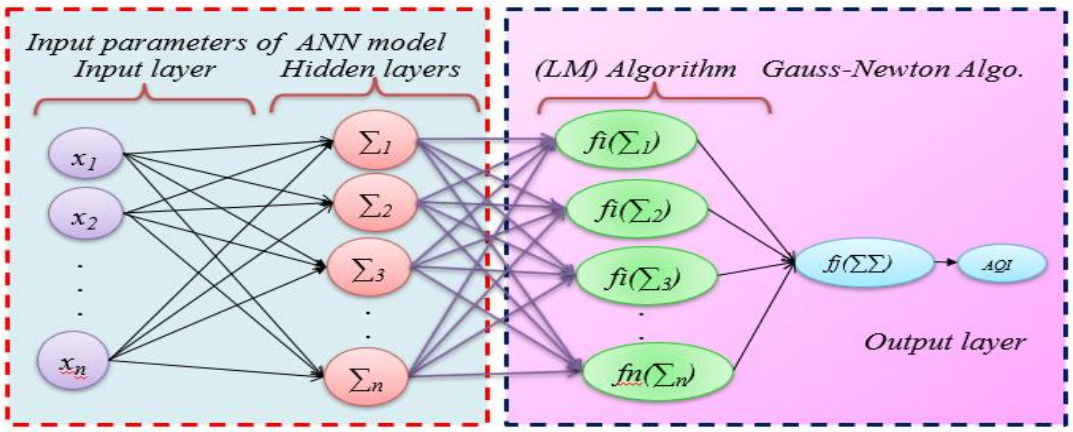

Figure 8. The architecture of Artificial Neural Network used for Air Quality Index Estimation

$$y_j = f_j(net_j) \tag{8}$$

where f_j is the activation function of neuron *j* and the net value *'net_j'* is the sum of weighted input nodes of neuron *j* which can be presented by Eq. (9).

$$net_j = \sum_{i=1}^{ni} w_{j,i}\, y_{j,i} + w_{j,o} \tag{9}$$

where, $y_{j,i}$ is the *i*th input node of neuron *j*, weighted by $w_{j,i}$, and $w_{j,o}$. When the training of the data set is completed, a high value of correlation coefficient decently describes that the data are highly correlated with the fit. It also shows that these parameters are significantly correlated, meaning that a change in one parameter will affect the other parameters. The histogram in Fig. 9 depicts the difference between the data values and the curve-fit. This figure also shows that the curve-fit errors are normally distributed.

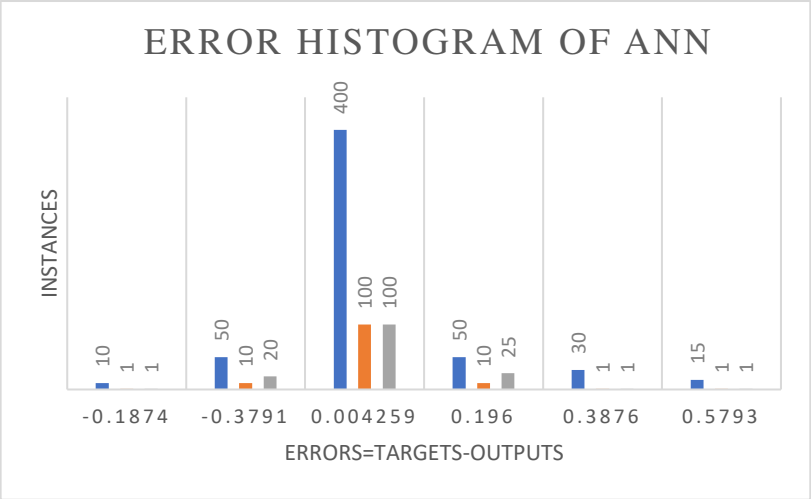

Fig. 9 ANNs Error Histogram for Training, Testing, and Validation

In this study, the redundant data were not used in the training process, as the ANN algorithm does not work well with redundant data. A multilayered perceptron (BPMLP) network with six inputs, eight processing units in the hidden layer, and one output parameter was considered for the training process. As ssen in fig.8, the back-propagation algorithms were used for training the network with LM tools' employment, which minimizes the divergence between the input and the output parameters. The outcomes predicted by the BPMLP algorithm were converted to air quality numerals that are recorded in Table 4. During the training process, it was found out that the solution had improved, the δ was decreased, the LM method approached the Gauss-Newton method, and the solution usually accelerated to the local minimum (Gavin, 2017). Sum square error (SSE) method was employed to assess the training process. The SSE for all training patterns and network outputs was computed using Eq. (10). The error rate is reasonable because redundant data and noisy data were excluded during the training, testing, and validation process. 60% of data was used for training, 20% for testing and 20% of data was used for validation. Excluding the outliers (the redundant data), the average absolute error was found 0.07147 %, and the sum of the squared errors was found 0.0251%.

$$E(x,w)=\frac{1}{2}\sum_{p=1}^{p}\sum_{m=1}^{M}e_{p,m}^{2} \tag{10}$$

where, as seen in Eq.10, $\boldsymbol{w}$ denotes the weight vector, and $\boldsymbol{e}_{p,m}$ refers the training error of the machine leaning algorithm.

Table 4. Air Quality Index vs outcomes of ANNs and ANFIS models for certain parameters

Sulphur Dioxide (SO_2, $\mu g/m^3$)	Carbon Monoxide (CO, $\mu g/m^3$)	Hydrogen sulfur, (H_2S $\mu g/m^3$)	Ozone (O_3, $\mu g/m^3$)	Nitrogen oxide (NO, $\mu g/m^3$)	Particular matters (PM_{10}, $\mu g/m^3$)	Air Quality Index, Observed	Air Quality Outcomes of ANNs	Air Quality Outcomes of ANFIS
12	4.4	339	75	4	60	198.70	197.7773	198.21
7	0.12	249	55	9	51	176.52	175.2971	176.733
4	0.12	164	57	9	49	155.87	154.7877	154.24
10	0.19	184	43	19	46	157.74	156.3981	157.49
11	0.29	338	49	20	53	159.03	158.9712	158.42
24	3.47	810	31	13	52	182.12	189.4121	184.78
31	0.64	887	29	24	47	145.91	144.7926	144.19
58	0.71	1020	35	16	67	98.58	99.06586	98.49
39	2.44	1198	37	15	49	73.96	73.71264	73.69
16	5.91	586	49	13	39	71.06	70.60034	70.90
9	4.37	88	43	23	40	71.24	70.76935	71.50
15	4.78	125	44	25	45	61.39	60.17372	60.25
19	5.25	216	38	27	59	97.16	97.38304	97.76
26	1.48	253	29	30	80	78.76	78.64256	78.28
19	3.99	314	52	17	37	77.64	77.6342	77.37
8	2.45	10	52	18	240	82.43	82.58674	83.03
10	7.71	19	43	19	109	91.02	91.56989	91.156
30	5.73	97	31	20	45	106.61	106.7636	106.63
24	3.47	810	31	13	52	108.50	108.7315	108.89
93	5.06	55	37	23	46	110.13	110.4067	110.33
67	2.07	88	39	17	54	127.85	127.4214	127.36
31	0.64	88	29	24	47	138. 68	137.7212	160.27
96	0.66	100	44	27	84	150.95	149.6483	149.34
29	4.31	106	34	13	69	156.02	154.6345	155.13

When using pattern *p,* as it is defined as in Eq. 11, *m* represents the index of outputs, from 1 to *M*, where *M* is the number of outputs.

$$e_{p,m} = d_{p,m} - o_{p,m} \tag{11}$$

'd' determines the desired output vector for air quality index (AQI), the actual output vector for AQI is represented by *'o'*. Considering the nodes and the links between the output node y_j of a hidden neuron j and network output o_m, a complex nonlinear relationship exists between the network parameters that can be defined simply by o_m and f_j, where o_m is the mth actual output of the network

representing the air quality.

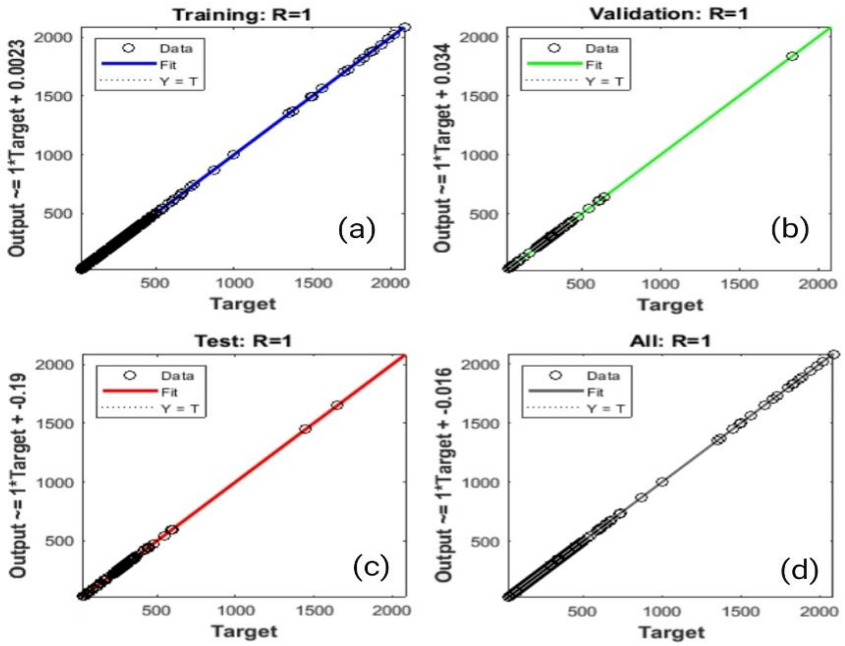

Figure 10. The regression plots of outputs for training, validation, and testing data

Fig. 10 (a), (b), (c), and (d) depict the targets of output tracks for training correlation coefficient (R), testing R and validation R. The value of R is close to 1 for training, and 0.91227 for validation of data. Similarly, the value of R for testing is 0.97948 and 0.98103 for validation. The training process was initiated as shown in fig. 7 (a), and the final training was carried out after several training steps and illustrated in Fig. 7 (b). The training, testing, and validation were converged at the three epochs with the validation performance of 92.3206. Thus, the result is acceptable since the final mean-square error and the absolute mean square errors are small, after several training steps, the error rates fell to 0.611236% and 0.080739%, respectively. It is also clear that the set errors of the training and testing have similar characteristics. For instance, no significant over-fitting has been obtained by iteration number thirteen, where the highest performance of the validation has occurred. On the other hand, ANN has a similar capability for the same data set of independent regressors' used for the ANFIS model training process. The training, testing, and validation of the ANFIS model were converged at the 60 epochs with the validation performance of 99.3206. The mean-square error and the absolute residual rate are small in this approach; after training, they fall to 0.611236% and 0.080739%, respectively. The errors of training and testing have similar characteristics. The low-level errors obtained were due to mainly insignificancy of over-fitting observed and occurred by iteration thirteen, where the best validation performance has been observed. Figure.11 (a) and (b) show the histogram of error distribution and the residual (c) of initial and final training stages of the machine

learning process, respectively. Hence, the prediction performance of this approach is high, and the average residual rate is low.

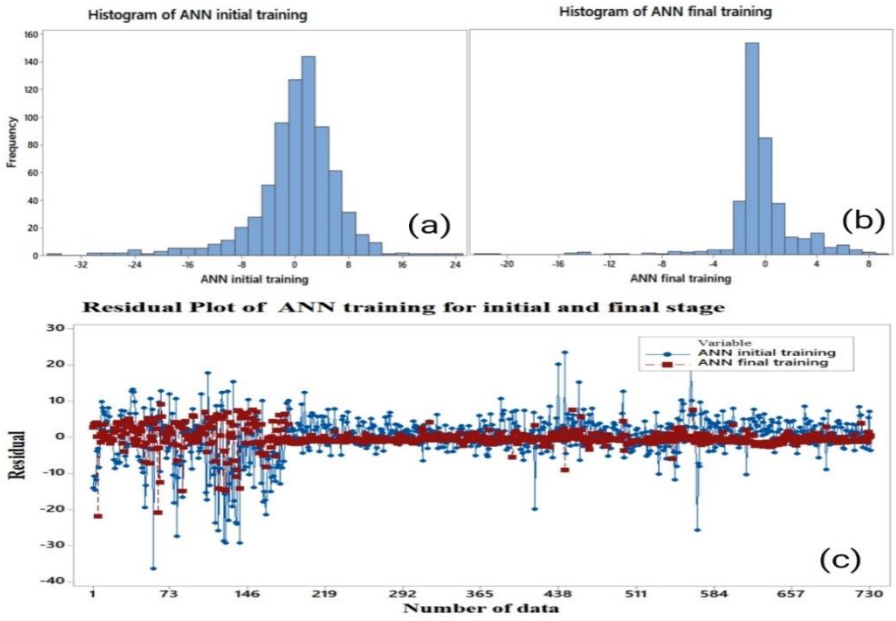

Figure 11. The training performance of machine learning and the residual of initial and final stages

4. Results and discussions

ANFIS and machine learning approaches are highly interrelated soft computing systems for information processing capable of deep learning. They were employed for the big-data advancement of the environmental systems, using the BPMLP algorithm and steepest descent approach to reduce the mean square error of the big data set training. The Levenberg-Marquardt (LM) approach was employed as an optimization method of ANNs, as a sub-technical machine learning approach for solving the pollutant parameters that have nonlinear relations. The results obtained were evaluated by fuzzy quality charts and compared with the US-EPA air quality standards statistically.

One of the most critical ecology issues is environmental pollution, including air, water, land pollution, etc. Emissions of sulphur dioxide and other pollutants are gradually rising as the number of industries grows up. Nitrogen oxides have been increasing in many locations. The widest spread of air pollution in these areas is mainly formed by the emissions created from the domestic industrial plants and transportation sources. Daily arithmetic averages of sulfur-dioxide, carbon-monoxide, hydrogen sulfite, ground-level ozone, nitrogen oxide, and particulate matter were collected from stations and used to model the air quality index.

Data accumulated over the last three years offered us a big data set which provided substantial deal for training the model to obtain an ANFIS model. The AQI of each pollutant was calculated by Eq. 1. and an air quality index was obtained for the cumulative effects of pollutants. Some gases are inert (like CO) and does not interact chemically with the others. However, we consider the relations statistically and mathematically. This data set was then employed to train the ANFIS and ANNs models to predict pollutants' air quality index. The degree level of inter-correlations between the pollutants shows that atmospheric pollution depends on various parameters, the relation of some pollutants with AQI are given in fig. 12. Ozone also has a negative correlation with AQI. There is a positive correlation between O3, SO_2, CO, NO_x and AQI. The associations between different air pollutants slightly vary in other relevant research that could be interpreted due to variations of different characteristics, such as location and unique meteorological factors. Table 5 shows the correlation matrix and multicollinearity between the pollutant parameters and their 'P' values. As the 'P' values are less than 0.05, is statistically significant.

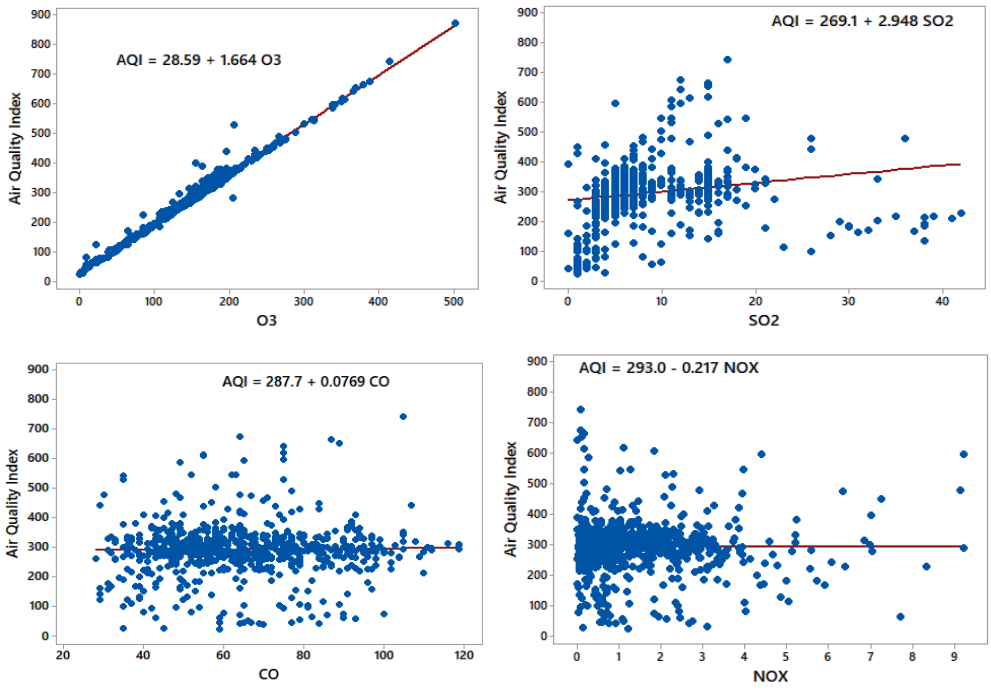

Figure 12. The association between O3, SO_2, CO, and NO_x with AQI.

Table 5. Correlation and Multicollinearity between the parameters and P-values

Environmental factors	AQI	Carbon monoxide	Hydrogen sulfite	Ozone	Nitrogen oxide	Particular matters
Sulfur dioxide	0.542					
P-value	0.000					
Carbon monoxide	0.142	0.145				
P-value	0.000	0.000				
Hydrogen sulphite	0.999	0.544	0.143			
P-value	0.000	0.000	0.000			
Ozone	-0.196	-0.288	-0.229	-0.21		
P-value	0.000	0.000	0.000	0.000		
Nitrogen oxide	0.137	0.205	0.131	0.140	-0.496	
P-value	0.000	0.000	0.000	0.000	0.000	
Particular matters	0.008	0.021	0.017	-0.034	-0.097	0.118
P-value	0.82	0.554	0.638	0.352	0.007	0.001

Sometimes, the forecast errors are computed in terms of percentages rather than amounts. Hence, in this study, the mean absolute percentage error (*MAPE*) was computed by finding the absolute error in each period, dividing this by the actual observed value for that period, and averaging these absolute percentage errors. The *MAPE* is a percentage and has no measurement units employed to calculate the accuracy of the same or different techniques on two entirely different series. Eq. 12 shows the *MAPE* calculation, and it is found 2.3747% for the AQI in this study.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|Y_t - \hat{Y}_t|}{|Y_t|} \quad (12)$$

On the other hand, the mean percentage error (*MPE*) was used to compute finding the error in each period. It is computed by finding the actual residual value for each period, then dividing by actual AQI values to obtain the % error, and at the end averaging these percentage errors. The *MPE* is calculated by Eq.13 and was found 0.3423% for this study, which is close to zero.

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{(Y_t - \hat{Y}_t)}{Y_t} \quad (13)$$

As a result, when *MAPE* of 2.3747% is compared to the *RMSE* of 5.64, the *MAPE* can be used to forecast the air quality data. A small *MPE* of 0.3423% reveals that the technique is not biased, while the value is close to zero, the techniques do not consistently over/or underestimate the AQI daily. The actual AQI observations versus the outcomes of ANN and ANFIS modeling approaches are given in fig. 13. The results clearly show that the outcomes of both models are close to the actual

AQI values and the air quality is good and moderate in Jeddah. There are some deviations during some periods this might be due to the effects of dust storms and particulate matters.

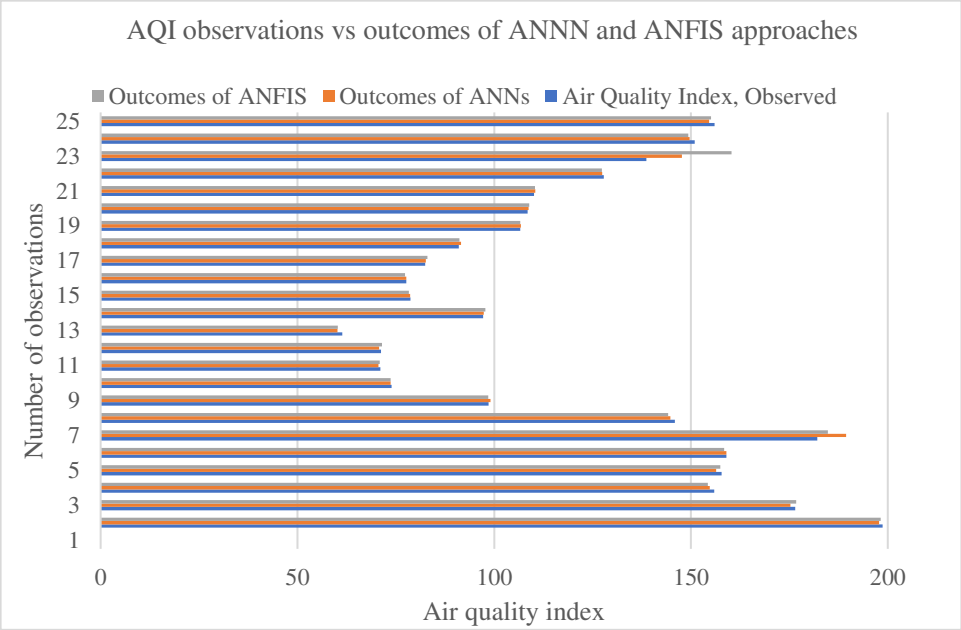

Figure 13. Air quality index observed vs the outcomes of ANN and ANFIS approaches

The ANFIS and ANNs model aims to construct an online and intelligent control strategy for air quality prediction. Both methods produced vigorous outcomes. Table 4 illustrates ANFIS and ANN models' outcomes for certain pollutants versus observed for air quality index. The average error of the ANFIS model was determined to be 0.10858 and which is 0.10362 for the ANN. The optimal number of rules was found six for the data set available. Moreover, the essential findings depicted that an additional number of membership functions and rules did not improve the ANFIS model's efficiency. Therefore, as it is given in fig. 4, six rules appear adequate to establish a rule based ANFIS model for AQI prediction. Fig. 5 depicts the fine-tuned MFs of pollutants; bell-shaped Gaussian MFs were employed for determining the membership degrees. The reason that the Gaussian MFs were employed is that the relations of parameters are nonlinear. Fig. 6 depicts the distribution of relative errors, determined for training and testing of the ANFIS model developed for this study. The ANFIS model outcomes for certain degrees of pollutants were given in Table 4, which provides the comparison of AQI obtained from the ANFIS model, and the observed AQI obtained from the US-EPA standard (US EPA, 1999). In this article, the back-propagation multilayer perceptron (BPMLP) algorithm was employed to perform nonlinear mapping of parameters. The BPMLP algorithm used the Levenberg-Marquardt (LM) approach as an optimization method for solving a nonlinear least-squares problem. Fig. 7 (a) and (b) show the initial and final training process, respectively. The

training process was successfully carried out because the mean-square error and the absolute mean square errors were low and were 0.611236% and 0.080739%, respectively. Similarly, Fig. 10 (a), (b), (c), and (d) show the training correlation coefficient (R), testing R and validation R; the R is 1 for training, 0.91227 for validation, and 0.97948 for the testing. ANN has a similar capability for the same data set of independent regressors' used for the ANFIS model training process. The low-level errors were obtained that are mainly because there is no significant over-fitting observed during iteration thirteen, where the best validation performance has been observed. Figure 11(a) and (b) show the histogram of error distribution and the residual (c) of initial and final training stages of ANN, respectively. A convergence was observed between the three parameters; hence the training process was ended.

Because the lack of identification of the cumulative effect of quality parameters in pollution issues, a novel trend has been inspired of combining randomness and fuzziness in evaluating environmental quality problem of air pollution in this work. Quality assessment in fuzzy sets expresses that the quality level of air is measured by membership degrees. The scatter plot of 100 principal component outcomes of AQI obtained for ANN and ANFIS models are illustrated in Fig. 14 (a) and (b), respectively.

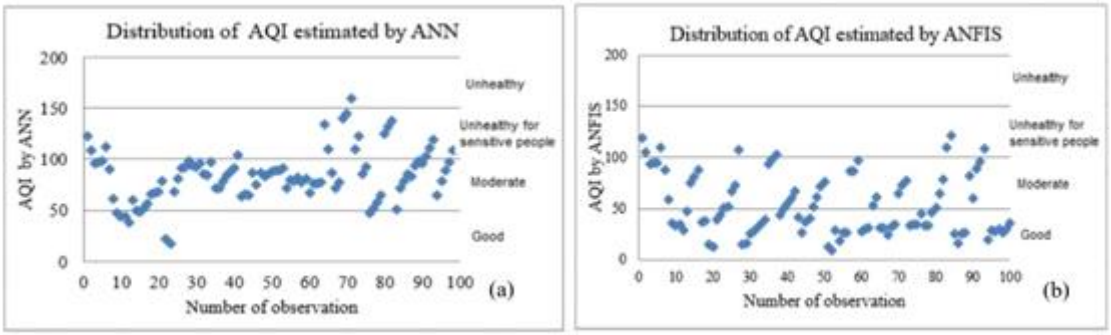

Figure 14. The air quality distribution and assessment by ANN and ANFIS models.

Figures 14 (a) and (b) show the outcomes of the AQI estimated by ANN and ANFIS models creating a category with the numerical values. The fuzzy quality charts with linguistic terms were employed along with the US environmental protection agency categories for air quality index (AQI) to evaluate the air quality in Jeddah. The ANFIS and ANN are more reliable and practical approaches to observe the air quality online, which add more flexibility than the crisp assessment of air quality offline. For an overall quality assessment, when the AQI is between 0 and 50, it is defined good air quality, if it

is between 51 to 100, the air quality is moderate. However, if it is above 100, the quality is poor and unhealthy; the sensitive groups are affected. Higher AQI creates hazardous (if it is above 300), which affects people's respiratory systems. EPA (US EPA, 1999) standards for air quality have been established to prevent several harmful effects of pollutants.

5. Conclusions

This study aims to envisage air quality and its distribution using AI techniques, such as neuro-fuzzy logic and ANNs as a machine learning approach. For the situation where the AQI values increase, people may encounter several symptoms of health concerns (US EPA, 1999). Air quality models' outcomes were found meaningful for warning the public earlier in case an unhealthy situation is encountered. The proposed methods in this work are practical, robust, and capable of estimating pollutants' cumulative effect inside the urban areas to reduce respiratory and cardiovascular mortalities. Consequently, the stability of air quality was correlated with the absolute air quality index. The findings show the remarkable performance of ANFIS and ANN-based Air Quality models for high dimensional data assessment.

Air pollution management involves capacity building, monitoring ground-based networks and systems for appropriate strategic and operational decision-making. Implementing these strategies requires quality controlling and assurance, modeling approaches, and institutional capabilities. Therefore, local, and global environmental policymakers can consider the presented methodologies and findings as a suitable, reliable, and useful technique in air quality assessment and management.

Data Availability Statement section

The data that support the findings of this study are openly available as mentioned in the reference section.

Author contributions

The individual contribution of the authors was as follows: O.T., A.S.K., and H.A. together designed research, provide extensive advice throughout the study reading to research designed, research methodology, data collection, assessment of the results and findings and revise the manuscript. A.B., M.A., and M.I. helped to draft, calculate, edit, and revise the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethical approval

All authors declare that there is no ethical violation in this manuscript. Also, this manuscript does not contain data belonging to others.

Consent to participate.

All authors have confirmed that this has not been published elsewhere and is currently not considered to be published elsewhere.

Consent to publish.

All authors agree that the article can be published in Environmental Science and Pollution Research.

References

Aggarwal, A. and Toshniwal, D., 2019. Detection of anomalous nitrogen dioxide (NO2) concentration in urban air of India using proximity and clustering methods. Journal of the Air & Waste Management Association, 69(7), pp.805-822.

Al-Alawi SM, Abdul-Wahab SA, Bakheit CS). Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone. Environ Model Software 2008;23:396–403.

Ansari, M.; Ehrampoush, M.H. Meteorological correlates and AirQ+ health risk assessment of ambient fine particulate matter in Tehran, Iran. Environ. Res. 2019, 170, 141–150.

Ayturan, Y.A.; Ayturan, Z.C.; Altun, H.O. **2018**, Air pollution modeling with deep learning: A review. Inter. J. Environ. Pollut. Environ. Model. 1, 58–62.

Alimissis, A., Philippopoulos, K., Tzanis, C.G., Deligiorgi, D., (2018) Spatial estimation of urban air pollution with the use of artificial neural network models. Atmos. Environ. 191, 205–213.

Bai, X.X., Dong, J., Rui, X.G., Wang, H.F. and Yin, W.J., International Business Machines Corp, 2017. Very short-term air pollution forecasting. U.S. Patent Application 14/939,522.

Biancofiore, F.; Busilacchio, M.; Verdecchia, M.; Tomassetti, B.; Aru_o, E.; Bianco, S.; Di Tommaso, S.; Colangeli, C.; Rosatelli, G.; Di Carlo, P., (**2017)** Recursive neural network model for analysis and forecast of PM10 and PM2.5. Atmos. Pollut. Res., 8, 652–659.

Christin, S, Hervet, É, Lecomte, N. Applications for deep learning in ecology. Methods Ecol Evol. 2019; 10: 1632– 1644. https://doi.org/10.1111/2041-210X.13256.

Cabaneros, S.M., Calautit, J.K., Hughes, B.R., (2019) A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Software, 119, 285–304.

Durante C, Cocchi M, Grandi M, Marchetti A, Bro R. Application of N-PLS to gas chromatographic and sensory data of traditional balsamic vinegars of Modena. Chemom Intelligent Lab Systems 2006;83:54–65.

Donnelly, A.; Misstear, B.; Broderick, B., (2015) Real-time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmos. Environ. 103, 53–65.

El Raey M. Air quality and atmospheric pollution in the Arab region, economic and social league of Arab States, Commission for Western Asia Joint Technical Secretariat of the Council of Arab Ministers Responsible for the Environment, University of Alexandria, Egypt, 2006.

EPA. Guideline on air quality models (revised). Research Triangle Park: NC: US Environmental Protection Agency; 40 CFR 51, 2005.

Fairbrass, A.J., Firman, M., Williams, C., Brostow, G.J., Titheridge, H. and Jones, K.E., 2019. CityNet—Deep learning tools for urban ecoacoustic assessment. Methods in Ecology and Evolution, 10(2), pp.186-197.

Gavin, HP., The Levenberg-Marquardt algorithm for nonlinear least-squares curve-fitting problems, Department of Civil and Environmental Engineering Duke University August 3, 2017.

Hanke JE, Wichern DW. Business forecasting, ninth edition, Pearson Int, 2009.

Ghoneim, O.A., Doreswamy; Manjunatha, B.R., (2017) Forecasting of ozone concentration in the smart city using deep learning. In Proceedings of the 2017 International Conference on Advances in Computing, Communications, and Informatics, ICACCI 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; V2017, 1320–1326.

Grivas, G.; Chaloulakou, A., (2006) Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. Atmos. Environ, 40, 1216–1229.

Hecht-Nielsen, R. (1988). Theory of the backpropagation neural network. Neural Networks, 1(Supplement-1), 445-448.

Hsu KJ. Time series analysis of the interdependence among air pollutants. Atmos Environ 1992;26:491–503.

Hvidtfeldt, U.A., Sorensen, M., Geels, C., Ketzel, M., Khan, J., Tjonneland, A., Overvad, K., Brandt, J., Raaschou-Nielsen, O., (2019) Long-term residential exposure to PM2.5, PM10, black carbon, NO2, and ozone and mortality in a Danish cohort. Environ. Int. 123, 265–272.

Ishibuchi H, Nakashima T, Murata T. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. IEEE Transact Systems Man Cybernetics-Part B Cybernetics 1999;29:601–618.

Iskandaryan, D.; Ramos, F.; Trilles, S. 2020, Air Quality Prediction in Smart Cities Using Machine Learning Technologies based on Sensor Data: A Review. Appl. Sci. 10, 2401.

Jang JSR, Sun CT, Mizutani E. Neuro-Fuzzy and Soft Computing, Prentice-Hall,1997.

Jorquera, H.; Perez, R.; Cipriano, A.; Espejo, A.; Victoria Letelier, M.; Acuna, G., (1998) Forecasting ozone daily maximum levels at Santiago, Chile. Atmos. Environ. 32, 3415–3424.

Kaur, G., Gao, J., Chiao, S., Lu, S., (2017). Air Quality Prediction: Big data and Machine Learning Approaches, Conference: 017 5th International Conference on Sustainable Environment and Agriculture (ICSEA 2017), Volume: 1, Los Angeles, USA.

675 Kassomenous PA, Kelessis A, Petrakakis M, Zoumakis N, Christidis TH, Paschalidou AK. Air
676 quality assessment in a heavily polluted urban Mediterranean environment through air quality
677 indices. Ecol Indic 2012;18:259–68.

678 Liu, F.; Chen, G.; Huo, W.; Wang, C.; Liu, S.; Li, N.; Mao, S.; Hou, Y.; Lu, Y.; Xiang, H.
679 Associations between long-term exposure to ambient air pollution and risk of type 2 diabetes
680 mellitus: A systematic review and meta-analysis. Environ. Pollut. 2019, 252, 1235–1245.
681
682 Maciąg, P.S., Kasabov, N., Kryszkiewicz, M., and Bembenik, R., 2019. Air pollution prediction with
683 clustering-based ensemble of evolving spiking neural networks and a case study for London area.
684 Environmental Modelling & Software, 118, pp.262-280.

685 Martens M, Naes T, Multivariate calibration. Chichester: Wiley, 1989.

686 Masmoudi, Sahar, et al. "A machine-learning framework for predicting multiple air pollutants'
687 concentrations via multi-target regression and feature selection." Science of The Total
688 Environment (2020): 136991.

689 Mitra S, Hayashi Y. Neuro-Fuzzy rule generation: a survey in soft computing framework. IEEE
690 Transact Neural Networks 2000;11:748–768.

691 Monks PS, Granier C, Fuzzi S, and et.al. Atmospheric composition change –global and regional air
692 quality. Atmospheric Environ 2009;43:5268–5350.

693 Munawar, S., Hamid, Dr ., Khan, M.S., Ahmed, A., & Hameed, N., (2017). Health Monitoring
694 Considering Air Quality Index Prediction Using Neuro-Fuzzy Inference Model: A Case Study of
695 Lahore, Pakistan. Journal of Basic & Applied Sciences. 12. 10.6000/1927-5129.2017.13.21.
696
697 Pal KS, Mitra S). Neuro-fuzzy Pattern Recognition: Methods in Soft Computing, John Wiley &
698 Sons, New York, 1999.

699 Pan, S., Choi, Y., Roy, A., and Jeon, W., 2017, Allocating emissions to 4 km and 1 km horizontal
700 spatial resolutions and its impact on simulated NOx and O3 in Houston, TX, Atmos. Environ.,164,
701 398-415.

702 Perez P. Prediction of sulfur dioxide concentrations at a site near downtown Santiago, Chile. Atmos.
703 Environ 2001;35:4929–4935.

Prasad, K., et. al., 2016. Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. Atmospheric Environment 128, 246-262.

Pawlak, I.; Jarosławski, J.; Pawlak, I.; Jarosławski, J., (2019) Forecasting of Surface Ozone Concentration by Using Artificial Neural Networks in Rural and Urban Areas in Central Poland. Atmosphere, 10, 52.

Reikard, G., (2019) Volcanic emissions and air pollution: Forecasts from time series models. Atmos. Environ. X, 1, 100001.

Rybarczyk, Y.; Zalakeviciute, R.; Rybarczyk, Y.; Zalakeviciute, R., (2018) Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. Appl. Sci., 8, 2570.

Rahman, M.M., Shafiullah, Md., Rahman, S.M., Khondaker, A.N., Amao, A., Zahir, Md., H.,(2020) Soft Computing Applications in Air Quality Modeling: Past, Present, and Future, Sustainability, Sustainability 2020, 12, 4045; doi:10.3390/su12104045.

Radojevi´c, D.; Antanasijevi´c, D.; Peri´c-Gruji´c, A.; Risti´c, M.; Pocajt, V. (**2019)** The significance of periodic parameters for ANN modeling of daily SO2 and NOx concentrations: A case study of Belgrade, Serbia. Atmos. Pollut. Res. 10, 621–628.

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A. and Jung, J., 2019. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. Neural Networks.

Silibello, C.; D'Allura, A.; Finardi, S.; Bolignano, A.; Sozzi, R. Application of bias adjustment techniques to improve air quality forecasts. Atmos. Pollut. Res. **2015**, 6, 928–938.

Sozzi, R.; Bolignano, A.; Ceradini, S.; Morelli, M.; Petenko, I.; Argentini, S., (**2017**) Quality control and gap-filling of PM10 daily mean concentrations with the best linear unbiased estimator. Environ. Monit. Assess. 189, 562.

Sharma, Ekta, et al. "A hybrid air quality early-warning framework: An hourly forecasting model with online sequential extreme learning machines and empirical mode decomposition algorithms." Science of the Total Environment 709 (2020): 135934.

736 Sowlat MH, Gharibi H, Yunesian M, Mahmoudi MT, Lotfi S. A novel, Fuzzy -based air quality
737 index (FAQI) for air quality assessment, Atmospheric Environ 2011;45:2050-2059.

738 Taylan, O., (2017) Modelling and analysis of ozone concentration by artificial intelligent techniques
739 for estimating air quality. Atmos. Environ. 150, 356–365.
740
741 Taylan O, Karagozoglu B. An Adaptive Neuro-fuzzy model for prediction of student's academic
742 performance. J Comput Industrial Eng 2009;57:732–741.

743 Taylan O. Estimating the quality of process yield by fuzzy sets and systems, Expert Systems Appl
744 2011;38:12599–12607.

745 Torney, C.J., Lloyd-Jones, D.J., Chevallier, M., Moyer, D.C., Maliti, H.T., Mwita, M., Kohi, E.M.
746 and Hopcraft, G.C., 2019. A comparison of deep learning and citizen science techniques for counting
747 wildlife in aerial survey images. Methods in Ecology and Evolution, 10(6), pp.779-787.

748 US EPA. Guideline for developing an ozone forecasting program. Environmental Protection
749 Agency, EPA-454/R-99-009, 1999.

750 Wahab AS, Bakheit SC, Al-Alawi S. Principal component and multiple regression analysis in
751 modeling of ground-level ozone and factors affecting its concentrations. Environ Model Software
752 2005;20:1263–1271.

753 Wahab SA. The role of meteorology on predicting SO2 concentrations around a refinery: A case
754 study from Oman. Ecologic Model 2006;197:13–20.

755 Wang, Deyun, et al. "A novel hybrid model for air quality index forecasting based on two-phase
756 decomposition technique and modified extreme learning machine." Science of The Total
757 Environment 580 (2017): 719-733.

758 WHO. Determination of Airborne Fibre Number Concentrations - A Recommended Method by
759 Phase Contrast Optical Microscopy (Membrane Filter Method). World Health Organization. Geneva,
760 2002.

761 Wilamowski, BM., and Yu, H. (2010) Improved Computation for Levenberg Marquardt Training,"
762 IEEE Trans. on Neural Networks, 21(6), 930-937.
763

764 Zhou, K.; Xie, R. 2020; Review of neural network models for air quality prediction. In Proceedings
765 of the Advances in Intelligent Systems and Computing AISC; Springer: Berlin/Heidelberg,
766 Germany, 1117, 83–90.
767 Zhou, Y., Chang, F.-J., Chang, L.-C., Kao, I.F., Wang, Y.S., **2019**, Explore a deep learning multi-
768 output neural network for regional multi-step-ahead air quality forecasts. J. Clean. Prod. 209, 134–
769 145.
770 Zhu, J.; Wu, P.; Chen, H.; Zhou, L.; Tao, Z.; Zhu, J.; Wu, P.; Chen, H.; Zhou, L.; Tao, Z. (**2018)** A
771 Hybrid Forecasting Approach to Air Quality Time Series Based on Endpoint Condition and
772 Combined Forecasting Model. Int. J. Environ. Res. Public Health, 15, 1941.
773
774
775

Figures

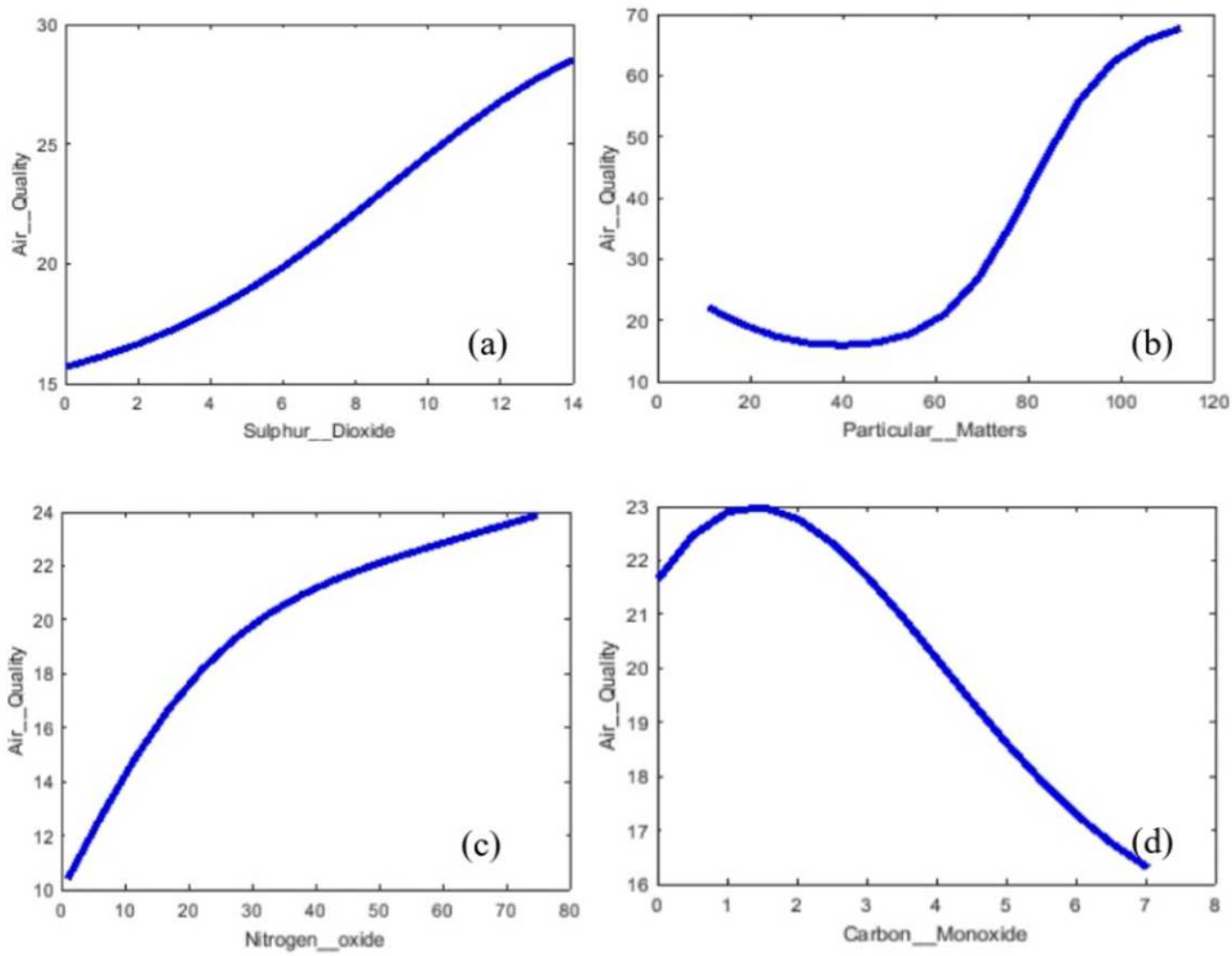

Figure 1

The impacts of pollutants on the air quality

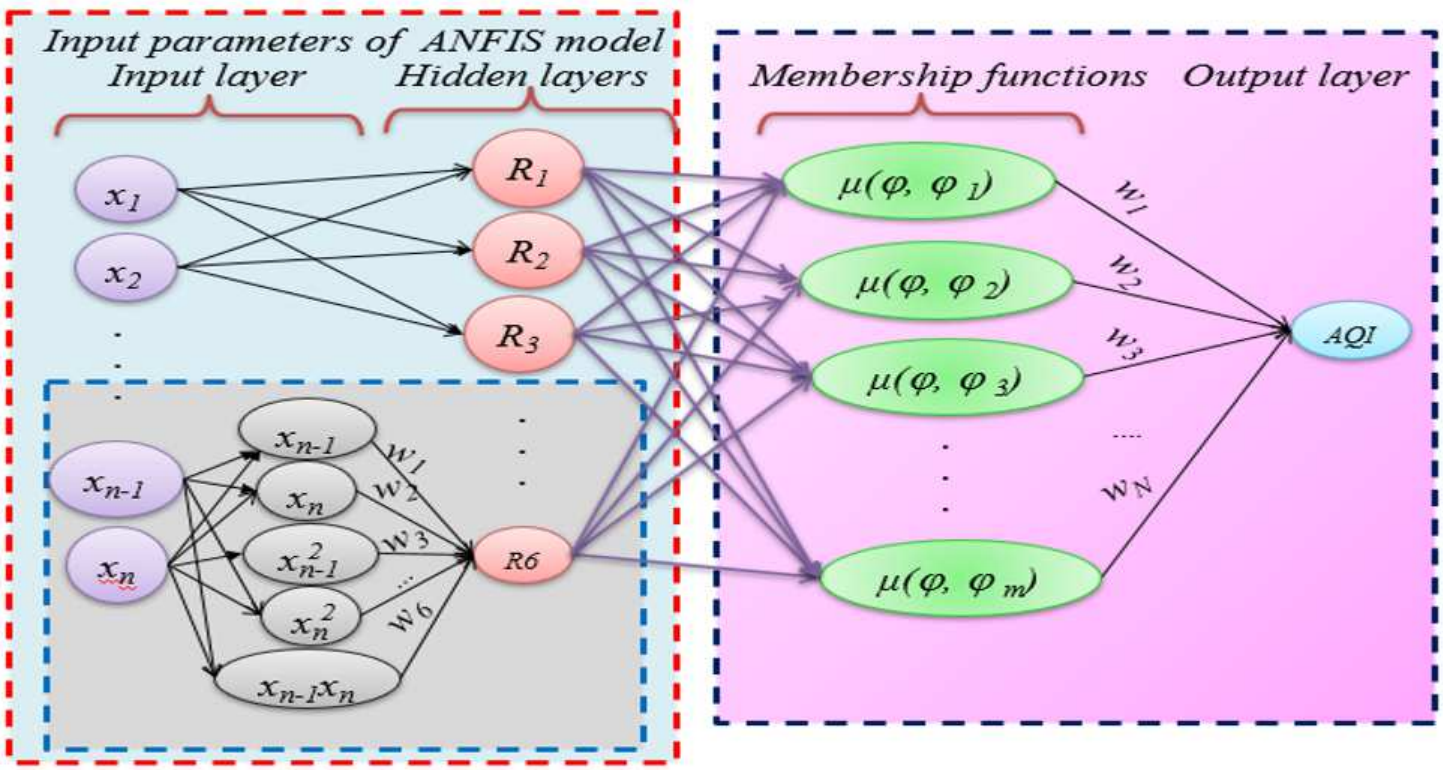

Figure 2

The ANFIS model architecture for air quality prediction

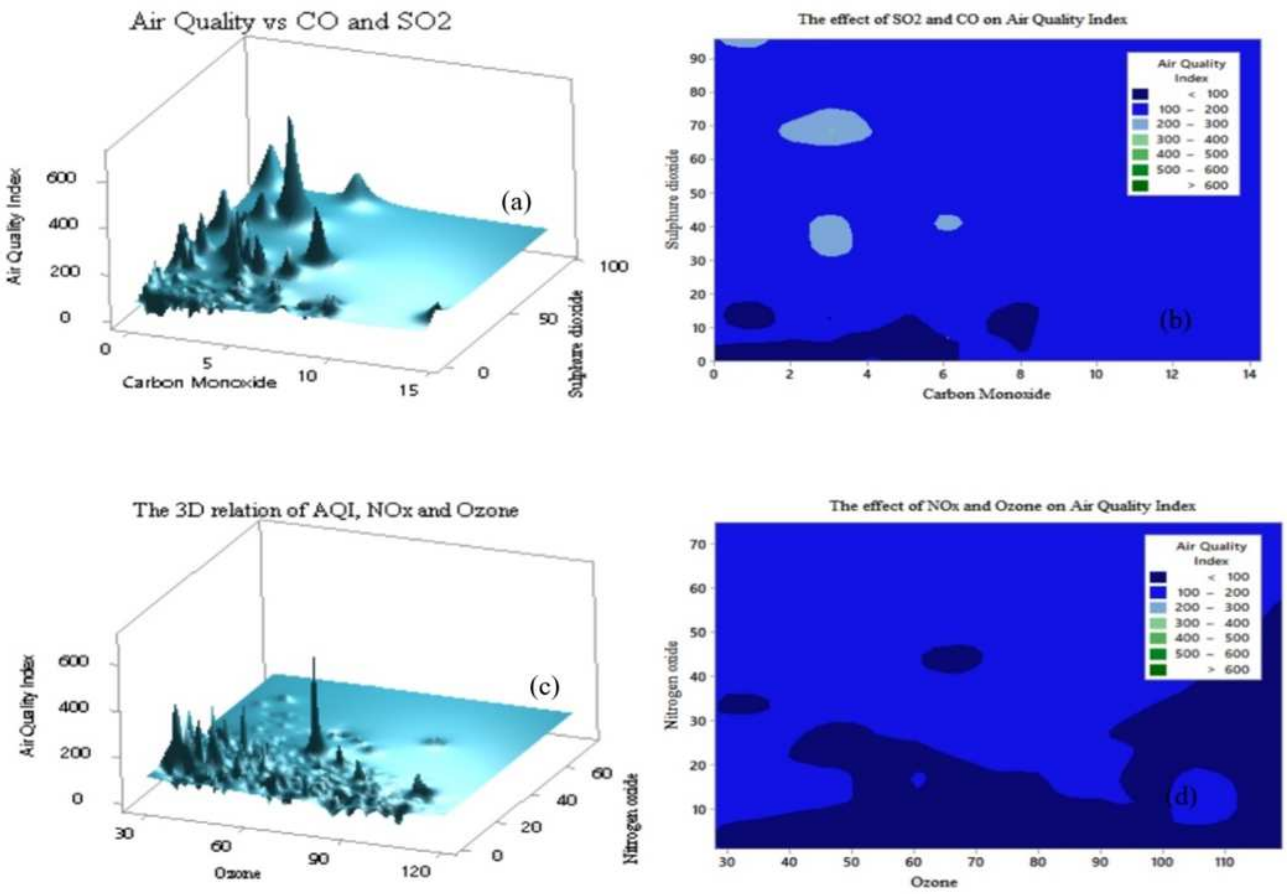

Figure 3

The impacts of pollutants on the air quality index

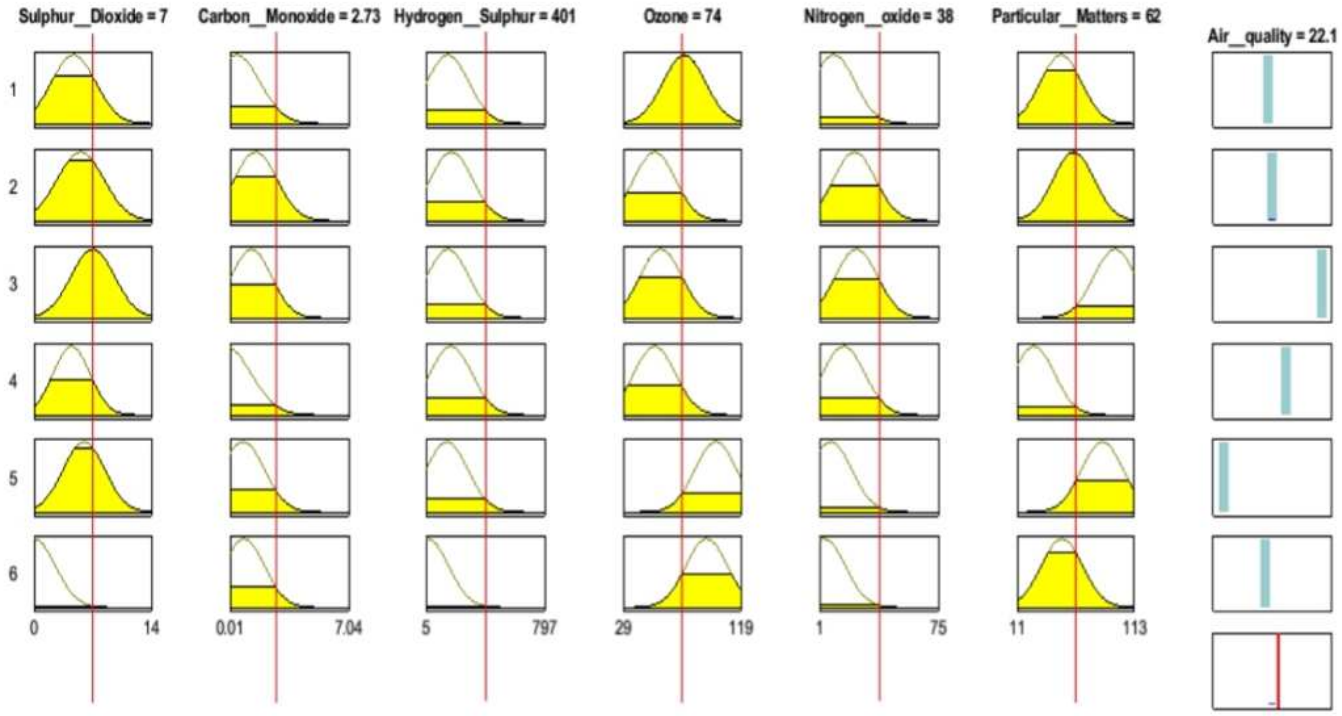

Figure 4

Fuzzy reasoning procedure for predicting air quality

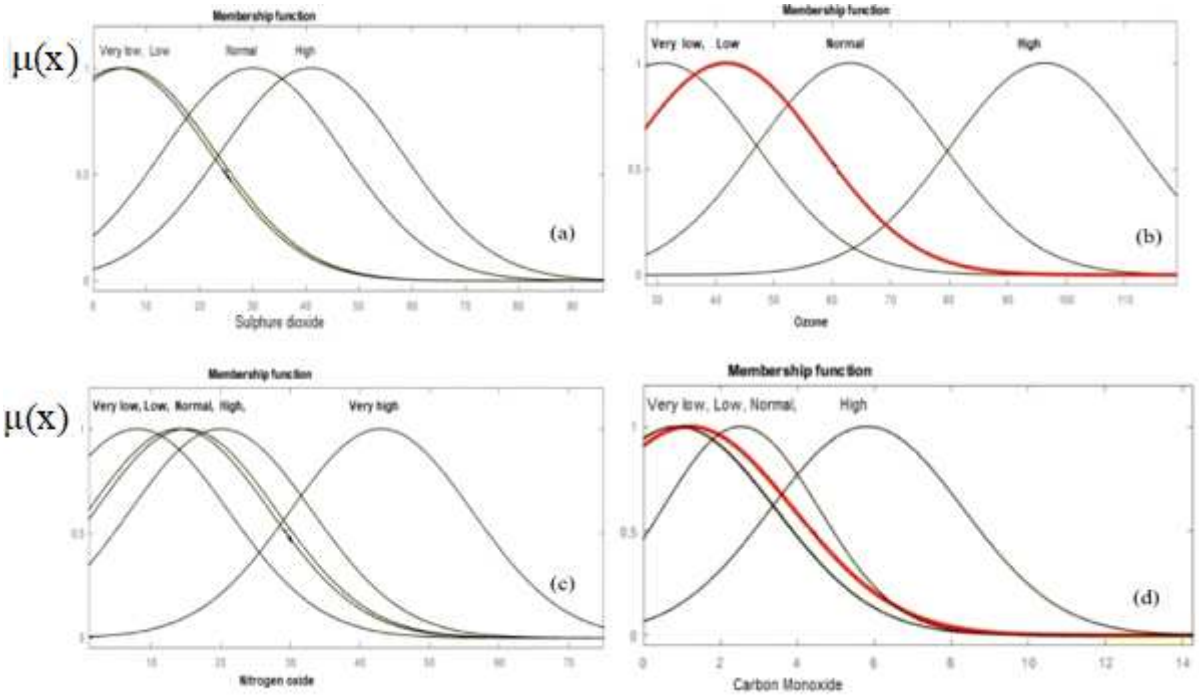

Figure 5

Membership functions and their terms set for air quality parameters.

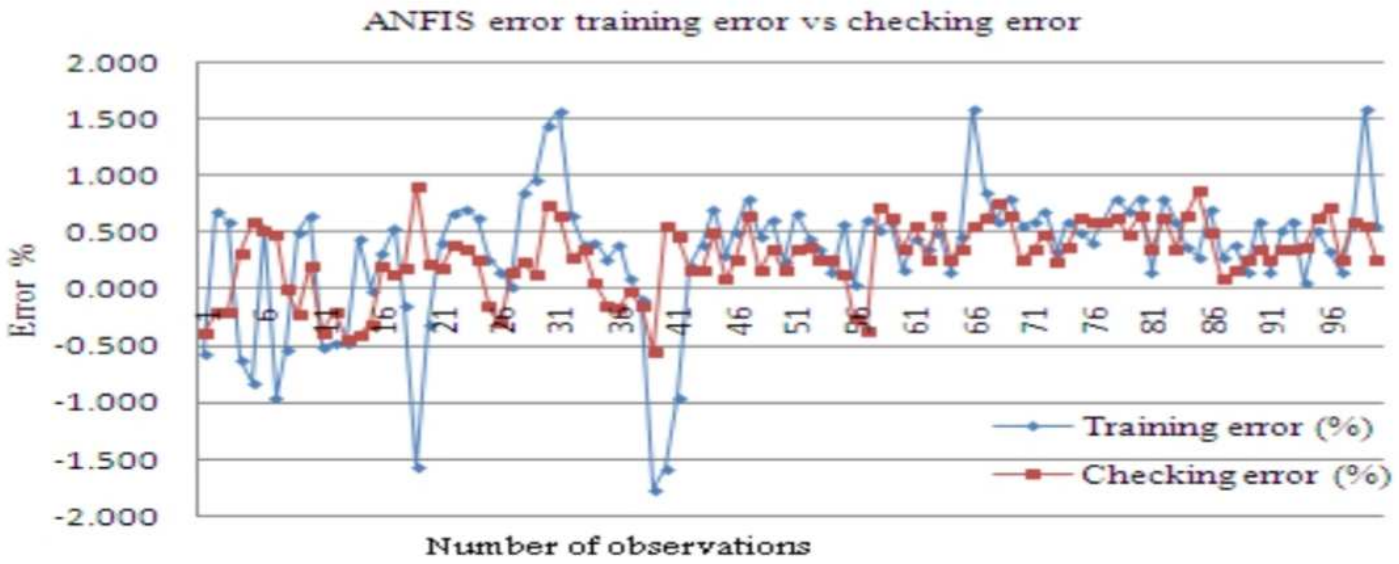

Figure 6

The training and checking error were determined for the ANFIS model.

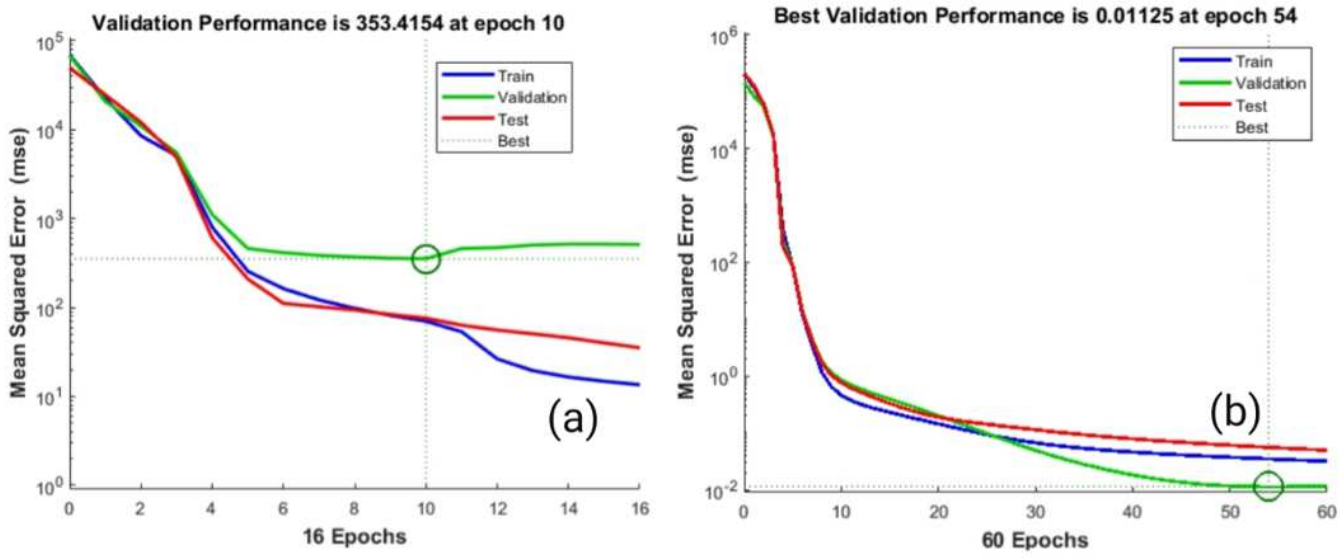

Figure 7

The weight distribution and allocation of training, testing, and validation for obtaining the minimum error.

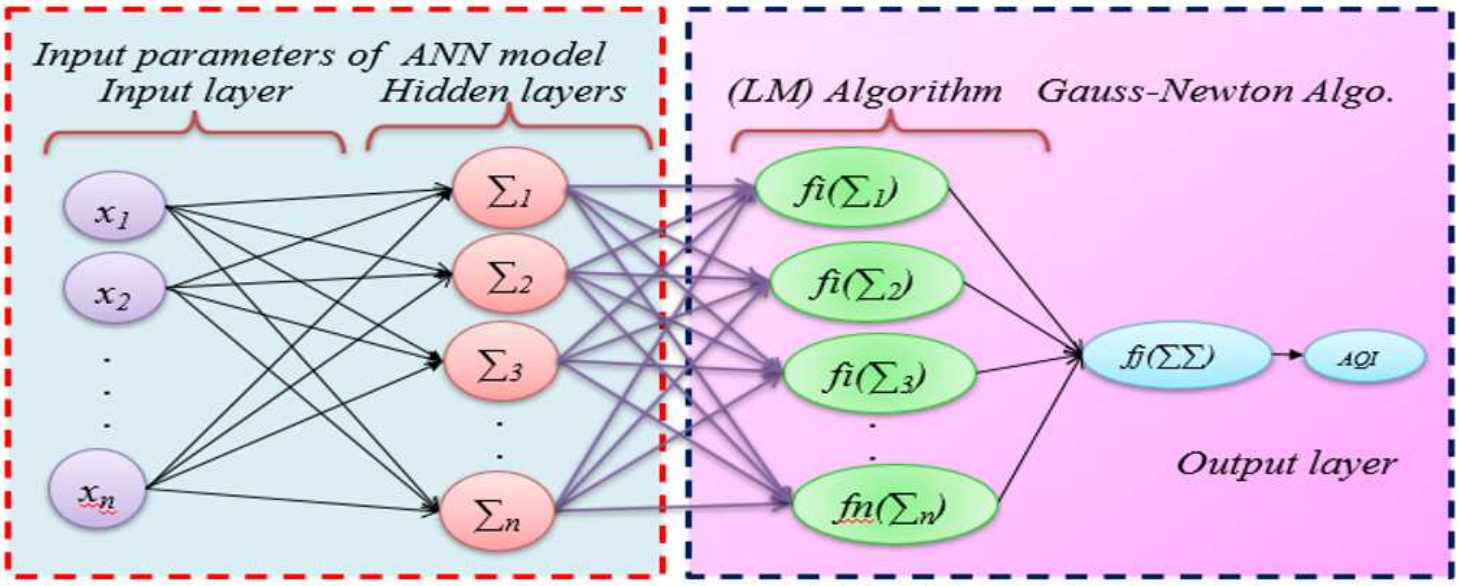

Figure 8

The architecture of Artificial Neural Network used for Air Quality Index Estimation

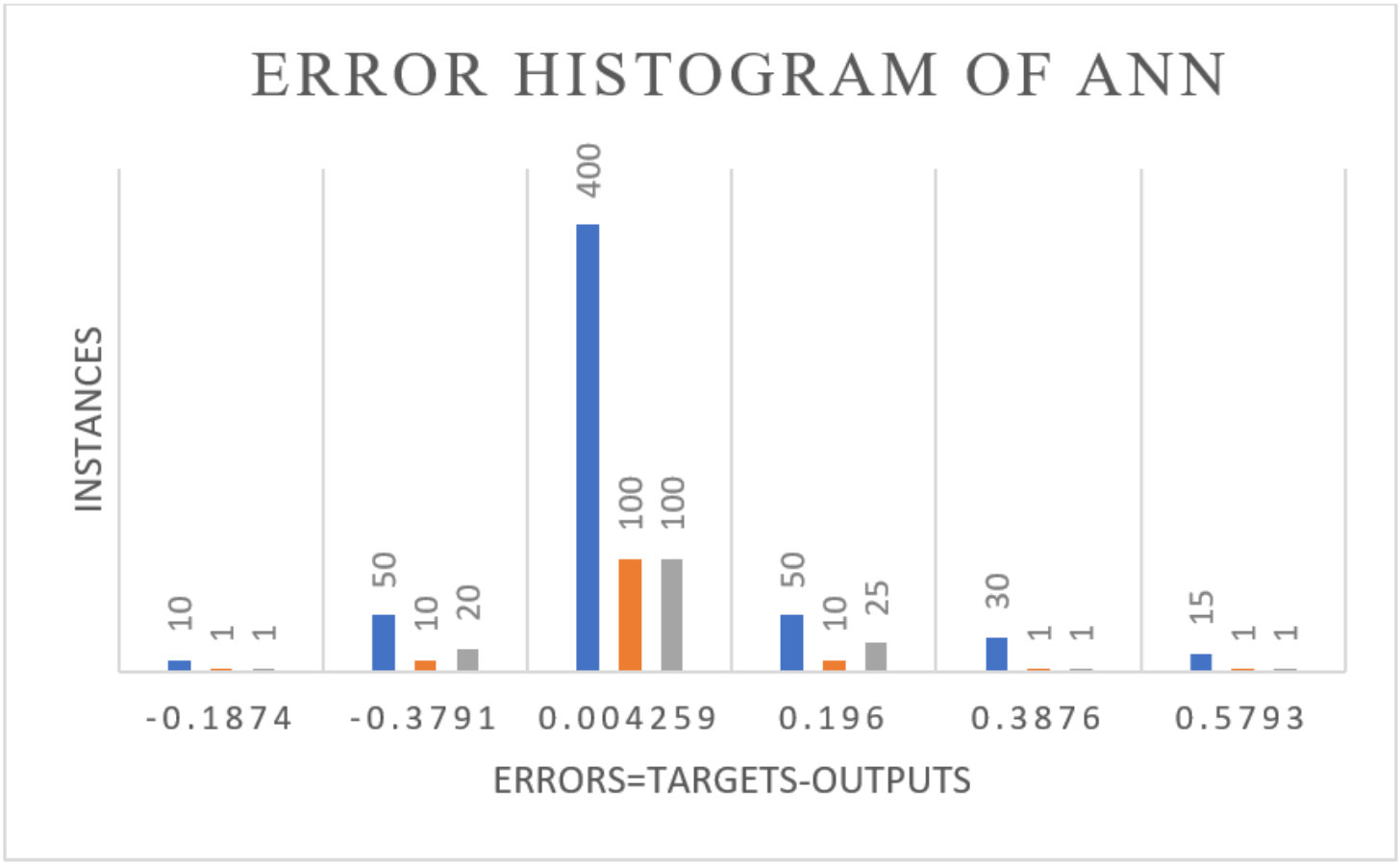

Figure 9

ANNs Error Histogram for Training, Testing, and Validation

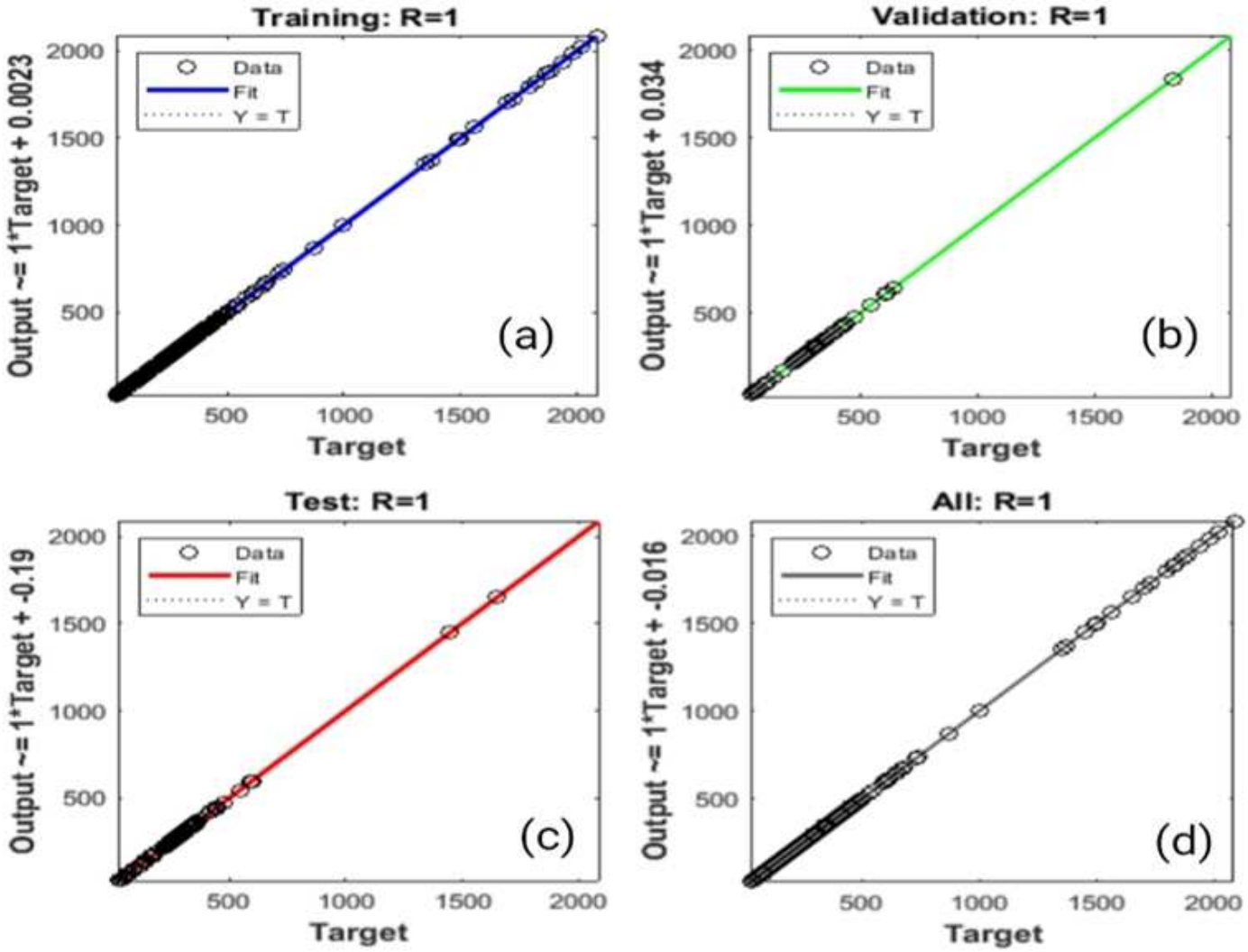

Figure 10

The regression plots of outputs for training, validation, and testing data

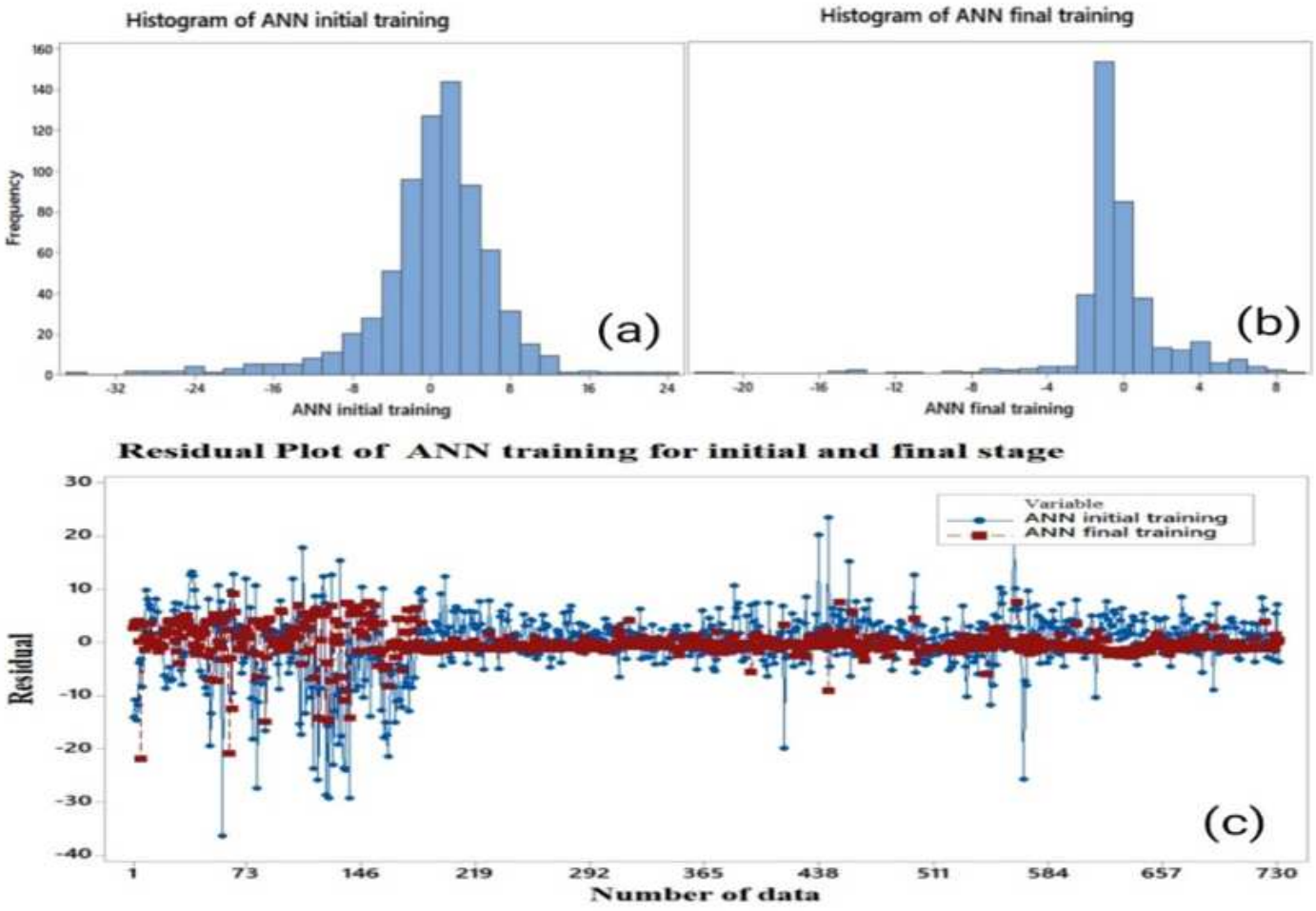

Figure 11

The training performance of machine learning and the residual of initial and final stages

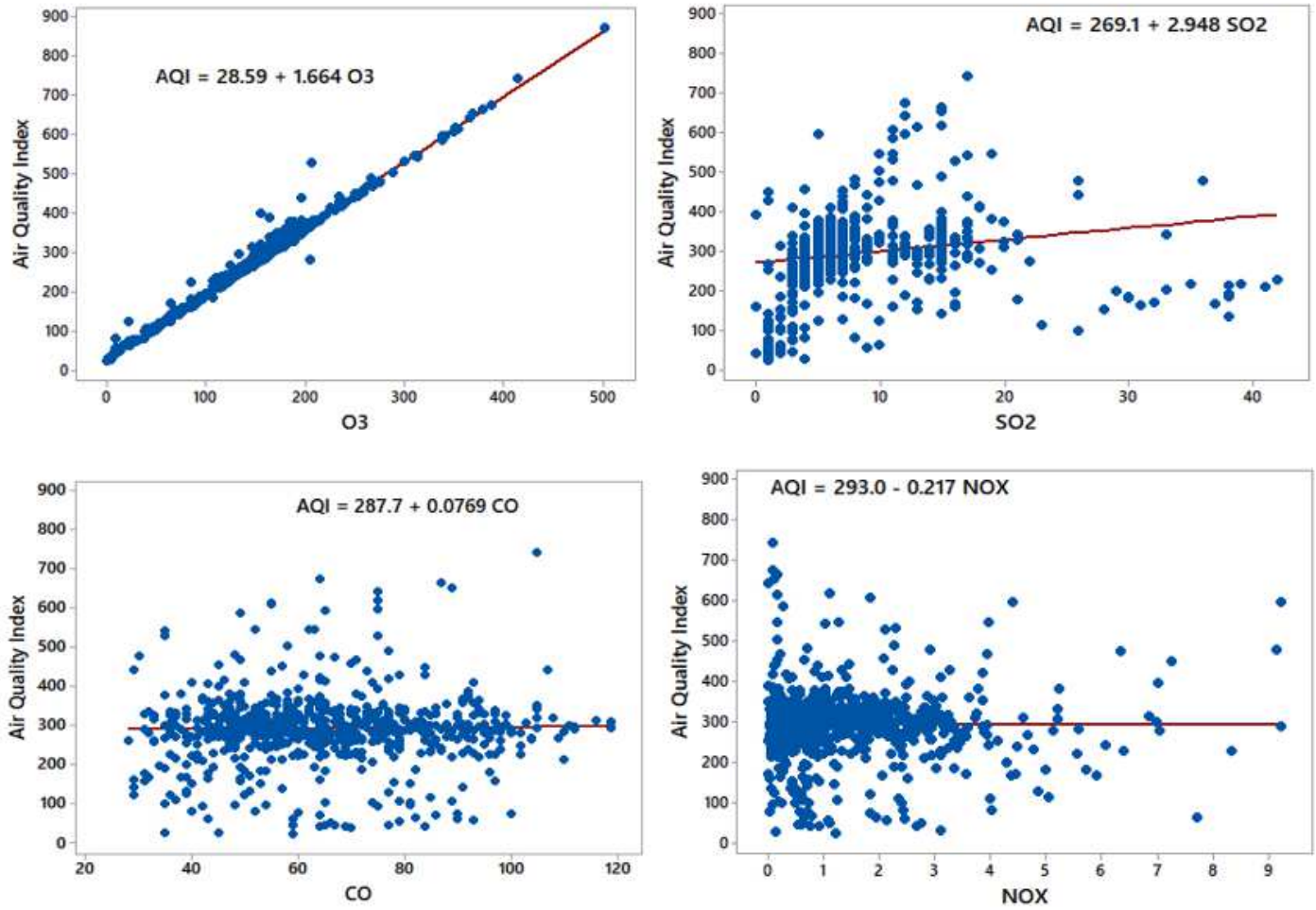

Figure 12

The association between O3, SO2, CO, and NOx with AQI.

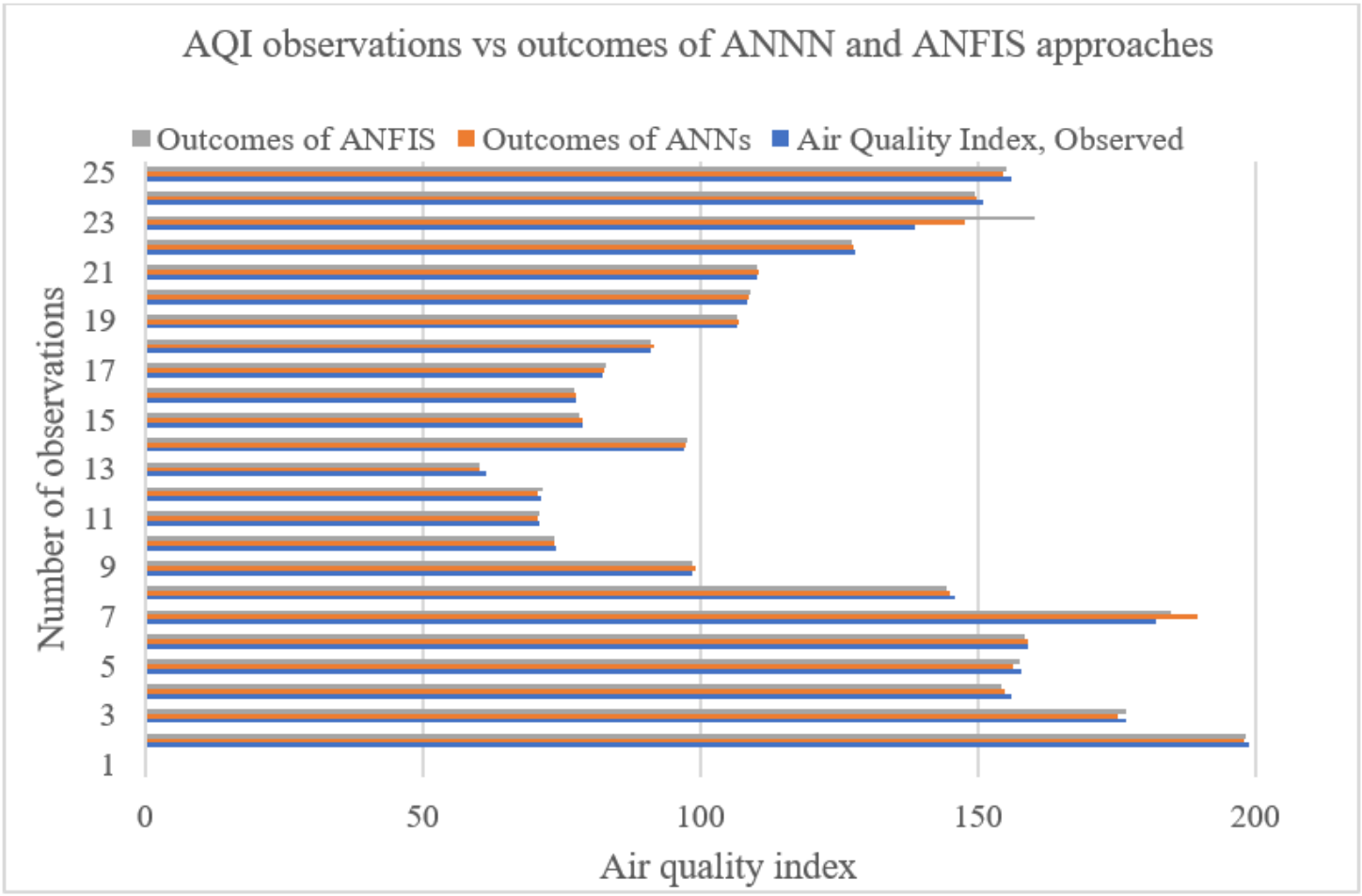

Figure 13

Air quality index observed vs the outcomes of ANN and ANFIS approaches

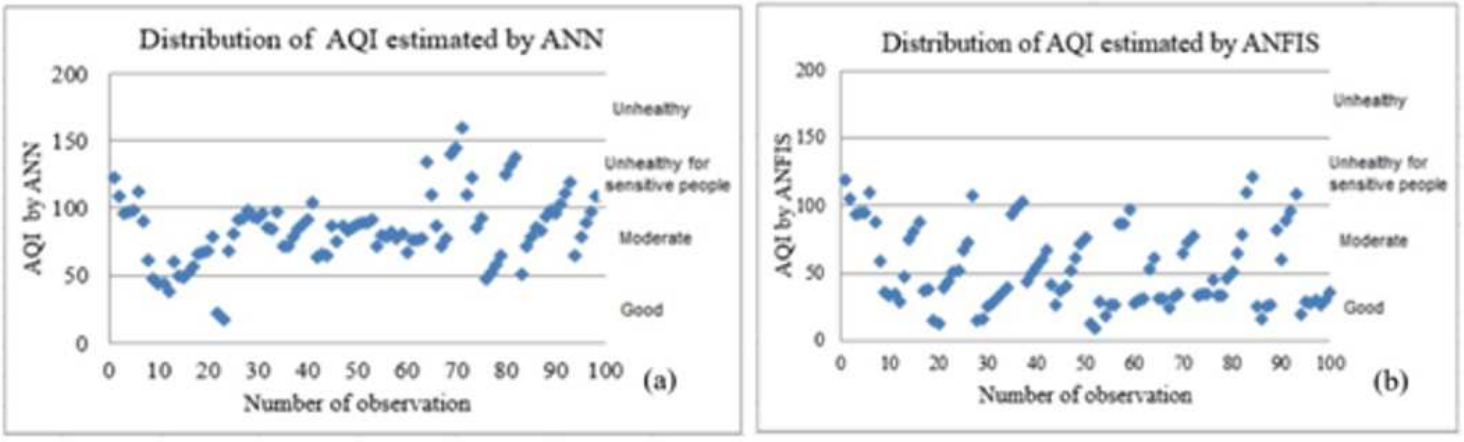

Figure 14

The air quality distribution and assessment by ANN and ANFIS models.