Health Monitoring Considering Air Quality Index Prediction Using Neuro Fuzzy Inference Model: A Case Study of Lahore, Pakistan

Saima Munawar1,2,*, Muhammad Hamid1, Muhammad Saleem Khan1, Ashfaq Ahmed2 and Noreen Hameed1,2

1School of Computer Science, National College of Business Administration & Economics, 40 - E/1, Gulberg III, Lahore, Punjab, Pakistan

2Department of Computer Science, Virtual University of Pakistan, 54, Lawrence Road, Lahore, Punjab, Pakistan

Abstract: For many years, improving air quality has been great attention of the whole world. It has been recognized that air pollution as a hypothetically hazardous type of environmental pollution and polluted air directly affects the human health. In Asian countries, it has converged less attention of ever growing most alarming and hazardous issue of air pollution. This paper presents a case study of Lahore city of Pakistan for the prediction of Air Quality Index (AQI) using hybrid approach of Neuro Fuzzy (NF) inference system. The ambient air data of Lahore was taken from the Environmental Protection Department (EPD) working under government of the Punjab. For results evaluation, data was recorded at different station in the period from April 2007 to May 2015. The fuzzy rules have been generated according to the Pakistan Environmental Protection Agency (PAK-EPA) standard of AQI. The NF Inference Model took the air pollutants such as Particulate Matter (PM2.5), Ozone (O3), Carbon Monoxide (CO), Sulphur Dioxide (SO2) and Nitrogen Dioxide (NO2) as inputs and predicted the air quality index as good, moderate, or unhealthy air. The results showed that NF based AQI prediction model classifies the AQI proficiently, robustly, and accurately as compared to conventional method.

Keywords: Air pollution, Fuzzy logic, Artificial Neural Network, Atmospheric Environment, Human Health effects, Air quality index.

1. INTRODUCTION

Air pollution directly affects the human health. Several common air pollutants such as O3, SO2, CO, NO2, and PM2.5, PM10 affect human’s health and due to this effect, human beings have been suffered in many diseases such as respiratory disorder, asthma, chronic obstructive pulmonary, cardiovascular and lungs cancer [1]. Several measurement methods were developed by various agencies such as United States Environmental Protection Agency (USEPA) [2], Japan International Cooperation Agency (JICA) [3], Pakistan Environmental Protection Agency (PAK-EPA) [4, 5]. In Asian countries, air pollution has converged less attention on air monitoring issues. PAK-EPA has investigated the air quality in the major cities of Pakistan, which was technically and financially funded by the JICA project. Air pollution of Lahore has raised thus supervising and forecasting the air pollution through air quality indexes are important due to their immediate impact on the health [6-10].

The AQI is a ranking scale for measuring and monitoring at a particular location of the air pollution status during a certain period of time such as 1hr, 8 hrs or 24 hrs. The main purpose of AQI is to implement some mandatory regulatory measures and report the ratio of air quality. When AQI is in peak condition, higher level of air pollution ratio affects the human health. Air pollutants has further divided in to four basic categories of pollutant ranges and quality of categories such as good, moderate, poor or hazardous [1, 11].

The hybrid approach of fuzzy logic and Artificial Neural network (ANN) has been used in air quality prediction. The advantages of this hybrid approach are that ANN has better learning ability, distributed knowledge representation, adaptation, parallel processing, and fault-tolerance, whereas the fuzzy logics strategy has dealt at a higher expert level with reasoning and it can interpret on the basis of rules [12]. The combination of these two approaches was used for different applications and classification tasks and it has gained benefit of both approaches [13]. The basic objective of this paper was to present hybrid approach that was a NF system for the prediction of AQI in the Lahore City. Five selective air pollutants were taken as the inputs to predict AQI on the basis of environmental hazard conditions that impact on human health by applying hybrid approach of fuzzy logic rules and neural network learning system. The experimental results have been analyzed on MATLAB with Adaptive Neural Fuzzy Inference System (ANFIS) simulation.
Five sections are presented in this paper. The significance of AQI has been described. Section 2 is an overview of NF inference system. Section 3 provided the detailed analysis of the design and structure of the proposed system. Section 4 provided experimental results of the work and discussed training and testing results. Section 5 presented conclusion and future direction of this domain.

2. BACKGROUND

In this section, the concept of ANFIS and conventional linear interpolation method for calculating AQI is explained.

2.1. Overview of ANFIS

ANN is a biologically inspired and mathematical model. It has nodes and units which are simple processing elements and interrelated layer wise to each other. The connections weights have stored in inter nodes processing network. The training examples have processed with learning ability of nodes and weights. Basically it has learned from these training examples or inputs processing. The output has determined through groups of inputs and connection weights. This model has been used over other approaches due to its ability of prediction, robustness to noisy data and to model generically and learning to non-linear systems [14].

The fuzzy logic system has basically emphasized on non-probabilistic uncertainty issues and it has less computational overhead to obtain the crisp output. The linguistic rules have been built according to each attribute with the help of membership function for inference engine, and output values have obtained through defuzzification techniques [15]. There are two common types of fuzzy inference system which has Mamdani model [16], Takagi and Sugeno model [17]. Takagi and Sugeno presented the fuzzy rule system in which, the output has determined of each rule and predefined function according to input variables such as

If first_input is MF1 and second_input is MF2 then output = a (first_input) +b (second_input) +c

MF1 and MF2 were the membership functions of the fuzzification /antecedent part and a, b, c were the adjustable parameters of a consequent part [17]. The learning of NF system modeling has based on ANN and their structure can be interpreted on the basis of fuzzy rules. Roger Jang has firstly proposed a hybrid ANFIS, it has same functionality as equivalent to FIS and it has used sugeno e tsukamoto type fuzzy model through hybrid learning algorithm. The basic architecture of ANFIS has based on six layers, the first layer has used as the input layer, inside four layers have used as hidden layers and last layer used as the output layer. All layers passed the forward signal to another layer and performed the individual task. The functionality of each layer has discussed in [13, 14].

The Jang presented the hybrid learning to train the ANFIS. The forward pass and backward pass have two passes in this network. Least square method has used in forward pass for identification of consequent parameters on the difuzzification layer. The gradient descent has used in the backward pass for updating premise parameters and errors has propagated backward [12].

2.2. Conventional Linear Interpolation Method for Calculating AQI

In this method, firstly it referred to all pollutants using breakpoints table that contain the values concentration, then calculate the index using linear interpolation method, it has used for calculating AQI [18]. The mathematical expression of this method has explained in [19]. This AQI method has not included combined effects of the major air pollutants and provided the overall assessment of air quality [20]. The AQI based on NF which can overcome the deficiencies of this benchmark method.

3. MATERIALS AND METHODS

This section provides an overview of the proposed NF based model for prediction of AQI. It has started from the overall analysis of design and structure of proposed model. Later on, it has demonstrated data analysis through a case study of Lahore. Then it has explained the feature selection and modeling of prediction. Finally, it has evaluated with the help of experiments and measures its performance. The details of each subsection have discussed below.

3.1. Design and Structure of Proposed Air Quality Prediction Model

The Basic structure of NF based air pollutant measure system has consisted of five sensors, which have used to monitor the air pollutants in the outside environment. The ultraviolet fluorescence method, gas phase chemiluminescence, non-dispersive UV absorption method, β ray absorption method, non-
dispersive infra-red (NDIR) sensors [21] have used to monitor and connected with the five fuzzifiers of the air pollutants which have SO2, NO2, CO, PM2.5, O3 of the fuzzy control system. This system has attained by NF hybrid learning algorithm for learning on the base of a fuzzy inference engine. The one difuzzifier output has connected through AQI actuator. The block diagram of proposed air quality prediction model as shown in Figure 1.

### 3.1.1. Case Study: Lahore City of Pakistan

Lahore is the second largest city of Pakistan and the capital city of Punjab province with approximate population 10,052,000 estimated in January 2015. The geographically, it is situated between 31°15′ to 31°45′ North and 74°01′ to 74°39′ East, it covered a total land area of 404 square kilometers and is still growing [22]. Lahore has a semi dry climate, June is the hottest month and monsoon season starts in July. January is the coolest month with dense fog. Approximately 75% of residents have their own conveyances in Lahore and on daily basis both public and private transportation has found. In Environmental statistics showed the key locations of Lahore where is deviating between WHO standards and local figures. In a 2003 report of Asian Development Bank showed that the severe situation of pollution exist in different parts of Lahore and it causes harmful effects on human health. Ambient air quality data includes different air pollutants such as Oxides of Nitrogen (NO2), Ozone (O3), Sulphur Dioxide (SO2), Suspended Particulate Matter (SPM), Lead Pb, Respirable Particulate Matter (PM10, PM2.5), and carbon monoxide with meteorological factors which directly affected on human health and also effected the agriculture, livestock. According to WHO’s recommended levels, Karachi and Lahore has exceeded air pollutants levels, especially carbon monoxide and particulate matter pollutants. In the past 20 years, sulfur dioxide has been 23-fold average increased in different emitting sectors such as power, industries and transport sectors and 25-fold nitrogen oxides has been increased in emitting power sector. Similarly the average of the fourfold carbon dioxide has been increased. The Pakistan’s per capita greenhouse gas (GHG) emissions have been found below the global average [7, 8, 23-25].

### 3.1.2. Dataset Selection and Air Pollutants Breakpoints

The air pollutants have been selected as an input which included 24 hours daily concentrations. The descriptions of these air pollutants are as follows: SO2

![Block diagram of proposed Air Quality Prediction Model](image-url)
is a reactive colorless gas and is produced such as coal and oil are burned when sulfur containing fuels. The major sources of this gas include refineries, power plants, and industrial boilers. PM$_{2.5}$ is the smallest 2.5 micrometer particles and major sources included vehicles, power plants, forest, fires, industrial and combustion processes. The concentration of ground level ozone has measured on the basis of 1 hour because when various pollutants emitted by sources forming near the ground such as industrial boilers, refineries, power plants, cars and chemical plants which chemically reacted in sunlight. O$_3$ pollution of Lahore is more likely to form during the hottest months such as May and June. Children, senior persons and people have been suffering in lung diseases such as bronchitis, asthma, emphysema and chronic are at higher risk from ground level ozone. The 8 hour concentration of CO is odorless and colorless gas. It has formed that carbon completely did not burn in fuels [26]. These pollutants are effecting the AQI by measuring daily wise air quality. AQI focused on human health which has affected after breath unhealthful or severe air for this reason neuro based system has proposed in the different scenario for measuring air pollution [27-29]. PAK-EPA has established for National air quality standards and protection of human health [21]. In this paper, the recorded daily wise average 24 hours air pollutants dataset of the selective station has used and it has provided by EPD through mobile and fixed station for the periods from 2007 to 2011 and 2014 to 2015, the sample set as shown in the appendix. This selection of the dataset has adopted due to the high ratio of air pollution has found in these periods as EPD remarked. The analysis of variance (ANOVA) test has implemented in the training and testing data set. It has tested the hypothesis that the means of these variable inputs have equal or significant. It has stated null and alternative hypothesis formulation as follows

Ho: Model has insignificant (on average of all variables are playing an insignificant role)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>2475285.513</td>
<td>4</td>
<td>618821.378</td>
<td>108.548</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>5672380.724</td>
<td>995</td>
<td>5700.885</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>8147666.237</td>
<td>999</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Groups</th>
<th>Sum of Squares</th>
<th>DF</th>
<th>Mean Square</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>2917260.375</td>
<td>4</td>
<td>729315.094</td>
<td>132.794</td>
<td>.000</td>
</tr>
<tr>
<td>Within Groups</td>
<td>2718576.964</td>
<td>495</td>
<td>5492.075</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>5635837.340</td>
<td>499</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ozone(O3) 1hrµg/m$^3$</th>
<th>Particulate Matter (PM2.5) 24 hrsµg/m$^3$</th>
<th>Carbon Monoxide (CO) 8hrs mg/m$^3$</th>
<th>Sulphur Dioxide (SO2) 24 hrsµg/m$^3$</th>
<th>Oxides of Nitrogen (NO2) 24 hrsµg/m$^3$</th>
<th>AQI Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-65</td>
<td>0-17.5</td>
<td>0-2.5 Favorable</td>
<td>0-60</td>
<td>0-40</td>
<td>Good</td>
</tr>
<tr>
<td>0-130</td>
<td>0-35</td>
<td>0-5 Moderate</td>
<td>0-120</td>
<td>0-80</td>
<td>Moderate</td>
</tr>
<tr>
<td>65-195</td>
<td>17.5-52.5</td>
<td>2.5-7.5 Unhealthy</td>
<td>60-180</td>
<td>40-120</td>
<td>Unhealthful</td>
</tr>
<tr>
<td>130-260</td>
<td>35-70</td>
<td>5-10 Hazardous</td>
<td>120-240</td>
<td>80-120</td>
<td>Severe</td>
</tr>
</tbody>
</table>

Table 1: ANOVA Test for Training Data

Table 2: ANOVA Test for Testing Data

Table 3: Build Breakpoints of Air Pollutants according to PAK-EPA Standards
The model has a significant role on average of all variables are playing a significant role. The significance level is set at $\alpha = 0.05$.

The test statistic to use is $F = \frac{s^2_b}{s^2_w}$ [30].

The ANOVA test for training data and testing data set calculated by using the SPSS [31] as given in Tables 1 and 2. The results showed that we rejected $H_0$ and it has also concluded that the model has significant.

### 3.1.3. Modeling: Prediction

The NF model has air pollutants as an input and one output of AQI prediction. The breakpoints of air pollutants have been mentioned in Table 3 and it has been built for a proposed model according to the standard of Pakistan National environmental quality ambient air. The structure of NF model of AQI measuring system has been designed as shown in Figure 2, which has six layers mapping in AQI, which has described below with each layer’s functionality.

In input layer 1, it transmitted the crisp input signal to the second hidden layer as Eq 1

$$a_{p,j} = O_{1,j}$$

//air pollutants (ap), O(output)

In layer 2 of hidden ANFIS architecture is also called fuzzification layer. The neurons of this layer...
represented antecedent part of fuzzy rules and it received the crisp input to determine the degree of belongingness of membership functions. Every neuron i represented the square neuron with membership function that is Eq 2

\[ O_{2,i} = \mu A_i \left( a_{p,i} \right) \]  

(2)

ap is input to neuron i and A_i is linguistic label such as favorable, moderate, unhealthy, hazardous and O_{2,i} has the membership function, where \( \mu \) represents the activation function of neuron i and has set to a triangular type of membership function. These were premise parameters. The triangular membership function equation has given in [13]. Which has three parameters (a, b, c) and x coordinates determined of three corners. Four triangular membership functions have been used for each air pollutants input such as favorable, moderate, unhealthy and hazardous which were used to show the ranges and lie on region 1, region 2 and region 3 of fuzzy variables.

In layer 3 of ANFIS architecture is second hidden layer and it is also called fuzzy rule layer. The single first order Sugeno fuzzy rule has used in each neuron, it received antecedent part of fuzzy rule from fuzzification layer, then multiplies the signal incoming and product output determined through rule strength firing, product operator has used to evaluate the aggregation of the antecedent parts as Eq 3

\[ O_{3,i} = w_i = \mu A_i \left( O_3 \right) \times \mu B_i \left( SO_2 \right) \times \mu C_i \left( PM_{2.5} \right) \times \mu D_i \left( NO_2 \right) \times \mu E_i \left( CO \right) \]  

(3)

\( i = 1,2,3,4 \)

Following are the fuzzy rules based on above membership function, Rule1: If (O3 is Favorable ranges) and (PM2.5 is Favorable ranges) and (SO2 is

---

**Figure 3:** Plot of air quality index prediction training error rate.

**Figure 4:** Plot of air quality prediction checking data error rate (above line shows the training error and below line shows checking data error).

**Figure 5:** Plot of AQI testing error rate
Favorable ranges) and (NO2 is Favorable ranges) and (CO is Favorable ranges) then (AQI is Good ranges), Rule2. If (O3 is Favorable ranges) and (PM2.5 is Favorable ranges) and (SO2 is Favorable ranges) and (NO2 is Favorable ranges) and (CO is Moderate ranges) then (AQI is Moderate output ranges), Rule3. If (O3 is Favorable ranges) and (PM2.5 is Favorable ranges) and (SO2 is Favorable ranges) and (NO2 is Favorable ranges) and (CO is Unhealthy ranges) then (AQI is Unhealthy ranges) //Responsible Pollutant is CO, so on ........

// Responsible Pollutant is SO2+NO2+CO

Rule1024. If (O3 is Hazardous ranges) and (PM2.5 is Hazardous ranges) and (SO2 is Hazardous ranges) and (NO2 is Hazardous ranges) and (CO is Hazardous ranges) then (AQI is severe ranges).

In layer 4, the third hidden layer is also called normalization layer. It received signals from third rule.

---

Figure 6: Surface graph between PM$_{2.5}$ and Ozone against the AQI output.

Figure 7: The simulated concentration and index value for air pollutants.

Figure 8: The comparison graph of Neuro Fuzzy and Linear Interpolation method.
layer and the ratio of normalized firing strength of the
given rules calculated by sum up of firing strengths of
all 1024 rules as Eq 4

\[ O_{4i} = \bar{w}_i = \frac{w_i}{w_1 + w_2 + w_3 + \ldots + w_{1024}} \]

In layer 5, the forth hidden layer is also called
difuzzification layer. In this layer, one diffuzzifier that is
AQI controlled by the actuator. It received the input from
both normalization layer neurons and initial input
signals and computed the consequent part of the given
rule based on sugeno type. This was a sugeno type
inference system, no allowed sharing rules on the
bases of output; all rules have different output
membership function in the base of input attributes.

In layer 6 is output layer and also called summation
layer. It computed the sum of all defuzzification output
neurons and overall ANFIS output produced by sugeno
type. To train the network in this system, it has selected
1000 epoch with 0.5 error tolerance rate. The total
rules have been built according to the triangular
membership functions of each air pollutant inputs that
are 4. So the total numbers of fuzzy rules have been
built \(4^4=1024\). Therefore the total number of parameters have been used in this ANFIS was 1084
which included 1024 of total number of linear
parameters and 60 were the nonlinear parameters.

4. RESULT AND DISCUSSION

The proposed model has designed for prediction of
AQI and has shown significant performance using NF
inference system with ANFIS MATLAB simulation as
compared to conventional linear interpolation method.
The 200 sample dataset from the environmental
protection department of Lahore has been used for
training, 100 pairs of dataset have been used for
testing and checking network. This hybrid learning
method has been used for training, testing and
checking with 1000 epochs and 0.5 error tolerance
rate. The training results of ANFIS are shown in Figure
3 with root mean square error (RMSE) rate 2.03 on the
1000 epochs.

The sample of 100 pairs of checking data set has
been used for the testing of the NF
inference system. The average error rate of testing the
system was 2.87 as shown in Figure 5. The surface
graph of two inputs PM\(_{2.5}\) and ozone against the AQI
output and the effect of these two inputs relation on
AQI as shown in Figure 6. After training, RMSE of
training, checking and testing data sets were 2.03,
0.14, 2.87 respectively and they were acceptable.

Some samples have been applied on NF based
model and linear interpolation for calculating AQI and
found that our model has given satisfactory simulation
results. The simulated concentration and index value
for air pollutants is shown in

5. CONCLUSION

NF system approach for prediction of AQI in the
Lahore City has been presented in this paper. Five
selective air pollutants have been taken as the inputs to
predict the AQI on the bases of environment hazards
conditions which impacts on human health by applying
hybrid approach. Intelligent air quality index prediction
system has shown significant performance with
learning proficiency of NF inference system. The
MATLAB ANFIS simulation results have shown that
this proposed design model will work efficiently in real
time environment for monitoring and predicting air
pollutants index in Lahore. Therefore, ANFIS can be
considered a useful architecture for attention by local
authorities for monitoring and assessment of AQI. It is
needed to build such system for analysis of daily basis
monitoring of public health.

6. FUTURE RECOMMENDATION

In future, it has suggested that this system will help
to design an advanced intelligent system with learning
capability to develop MANFIS and CANFIS architecture
of whole system. Microelectronics technology can also
be used to develop FPGAs control chip for monitoring
air pollution in the environment.

ACKNOWLEDGEMENTS

This work was guided by School of Computer
Science NCBA&E, Lahore. The authors also thank to
the Environmental Protection Department (EPD) in
Lahore for providing the ambient Air data.
APPENDIX

Sample of Air Quality Dataset of Lahore Regions (2014-2015) has taken from the Environmental Protection Department (EPD) working under the Government of the Punjab as shown in Table 4.

Table 4: Sample of Air Quality Dataset of Lahore Regions

<table>
<thead>
<tr>
<th>Station</th>
<th>Component</th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
<th>(Max-)Time</th>
<th>Unit</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>PJB-F2</td>
<td>CO</td>
<td>0.456</td>
<td>1.856</td>
<td>5.613</td>
<td>8/5/2015 22:30</td>
<td>mg/m3</td>
<td>102.71%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>MC</td>
<td>58</td>
<td>133.4</td>
<td>246</td>
<td>8/5/2015 9:40</td>
<td>ug/m3</td>
<td>66.46%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>NO</td>
<td>9.5</td>
<td>9.8</td>
<td>10.1</td>
<td>8/5/2015 22:40</td>
<td>ug/m3</td>
<td>101.70%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>NO2</td>
<td>33.5</td>
<td>34.2</td>
<td>34.2</td>
<td>8/5/2015 18:30</td>
<td>ug/m3</td>
<td>101.70%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>NOx</td>
<td>25.5</td>
<td>25.8</td>
<td>26.1</td>
<td>8/5/2015 6:20</td>
<td>ppb</td>
<td>101.70%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>RH</td>
<td>37.9</td>
<td>39.8</td>
<td>41.6</td>
<td>8/5/2015 22:40</td>
<td>%</td>
<td>4.3%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>Radiation</td>
<td>2.8</td>
<td>2.9</td>
<td>3.2</td>
<td>8/5/2015 23:00</td>
<td>w/m2</td>
<td>4.3%</td>
</tr>
<tr>
<td>PJB-F2</td>
<td>SO2</td>
<td>32.5</td>
<td>33</td>
<td>33.5</td>
<td>8/5/2015 13:50</td>
<td>ug/m3</td>
<td>102.71%</td>
</tr>
</tbody>
</table>

REFERENCES


[23] Chandrappa R, Kulshreshtha UC. Safety Issues in Sustainable Air Pollution Management. In Sustainable Air
https://doi.org/10.1007/978-3-319-21596-9_9


[28] Lei KS, Wan F. Pre-processing for missing data: a hybrid approach to air pollution prediction in Macau. In Automation and Logistics (ICAL), 2010 IEEE International Conference on 2010; (pp. 418-422). IEEE.


Received on 26-11-2016 Accepted on 27-03-2017 Published on 12-04-2017

https://doi.org/10.6000/1927-5129.2017.13.21

© 2017 Munawar et al.; Licensee Lifescience Global. This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.