Air Quality Prediction and Analysis using ANN and ANFIS

Tarun Kumar Ravipati (email: travipati1@student.gsu.edu)
Sudheer Nimmagadda (email: snimmagadda2@student.gsu.edu)

Abstract – **Pollutants in the atmosphere is a major problem in metropolitan cities due to vehicle and factory pollution. There is a need to take preliminary measures to overcome this problem since it leads to many health related issues like asthma and heart problems. So, there is a need to predict the level of these toxicants during various times of the day in order to warn the people suffering from these diseases so that they can be more cautious during the critical intervals of the day. We have identified the major pollutants of the atmosphere as NO_2 and O_3 along with 11 other pollutants which are either meteorological or geographic and particulate concentrations. We have used the techniques for prediction such as Artificial Neural Networks(ANN) and ANFIS(Adaptive Neuro Fuzzy Inference Systems) and compared them to find out that one perfect method that can precisely predict the concentrations for every 8 hours of the day.**

Index Terms— **ANFIS, ANN, PCA-ANFIS, PCA-ANN, Regression Analysis**

I. Introduction

THIS paper uses various convenient, conventional and efficient methods to predict the pollutants of the atmosphere at a particular period of a day. These techniques are used to generate alerts when the toxicant levels are beyond the permissible value so that the individuals can be warned.

The motivation behind this came from the amount of serious damage that the pollution in major cities has been causing damage to people living in India. We have identified Delhi as the city that has been most effected by pollution. The major pollutants that are considered responsible for this are identified to be NO_2 and O_3. The threshold values for the safe existence of these pollutants provided by the Center for Pollution Control Board(CPCB) of India are 200 µg/m3 and 100 µg/m3 for NO_2 and O_3 respectively. In order to prevent the effects of these pollutants, there is a need in the local and global communities for efficient air prediction models. These models predict the pollutant concentrations based on background concentration of pollutants, meteorological and geographical conditions owing to other local characteristics. The detailed studies of various models are discussed in this paper.

II. Data Collection and Preprocessing

A. Data Collection

The data set that is used for training, testing and validation of the data is collected from the official website of Center for Pollution Control Board(CPCB), www.cpcb.gov.in. The prediction of the pollutant levels in the future is obtained by considering the above dataset. There are multiple datasets that are available but we consider the region Delhi and an area near the airport where people reported optimum levels of pollution.

B. Data Analysis

The data set that is collected consists of many parameters of which a set of 13 parameters that cause the maximum level of pollution to the atmosphere consisting of both particulate concentrations and meteorological phenomenon. The Data is collected for a span of every 8 hours from 2011-2015. The Analysis of the data is classified as follows.

Training Data: 2011-13 Dataset.
Testing Data: 2014 Dataset.
Validation Data: 2015 Dataset.
The list of parameters that are considered are-

1. Particulate Concentrations:
   a. Benzene (BEN)
   b. Carbon Monoxide (CO)
   c. m,p-Xylene (MPXY)
   d. Nitrous Oxide (NO)
   e. Nitrogen Dioxide (NO_2)
   f. Ozone (O_3)
   g. Sulphur Dioxide (SO_2)
   h. Toluene (TOL)
2. Meteorological Concentrations:
   a. Bar Pressure (BP)
   b. Relative Humidity (RH)
   c. Temperature (TEMP)
   d. Wind Direction (WD)
   e. Wind Speed (WS)

C. Data Preprocessing

The raw data that is collected consists of a modest amount of inconsistencies in the form of unavailable and extreme/unrealistic observations. For example, there can exist negative values for the temperatures because of the magnetic repulsions on the temperature sensor due to various factors. This type of data that is obtained is not reliable and cannot be used for training and testing the data. So, we need to polish the

Tarun Kumar Ravipati, is with the Georgia State University, Department of Computer Science Atlanta, GA-30324, USA (email: travipati1@student.gsu.edu)
Sudheer Nimmagadda, is with the Georgia State University, Department of Computer Science Atlanta, GA-30324, USA (e-mail: snimmagadda2@student.gsu.edu).

data in order to obtain a sequence of normalized values that are bound by a unique range of values.

The unrealistic data is cleaned and merged in comparison with the existing values. Sometimes, the dataset consists of missing values in certain attributes. This can be eliminated using the nearest neighbor method i.e, the end points of the gaps were used as estimates for all the missing values between them. The absolute values of the negative values in each parameter are considered. Finally, the standardization is done where the features are shifted by mean and scaled by their variance to obtain a standard normal distribution. This method of standardization is called *Normalization*. The result of normalizing the data i.e, the preprocessed data of particulate and meteorological concentrations of BEN, BP and NO_2, O_3 are represented in the form of box plots in Figure 1 and Figure 2 respectively.

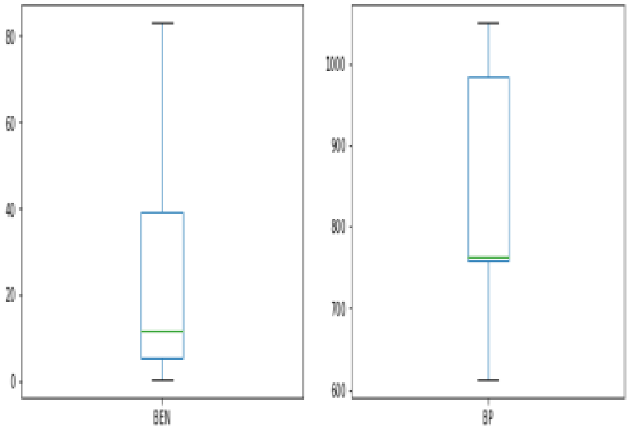


Figure 1: Box Plots for BEN and BP parameters

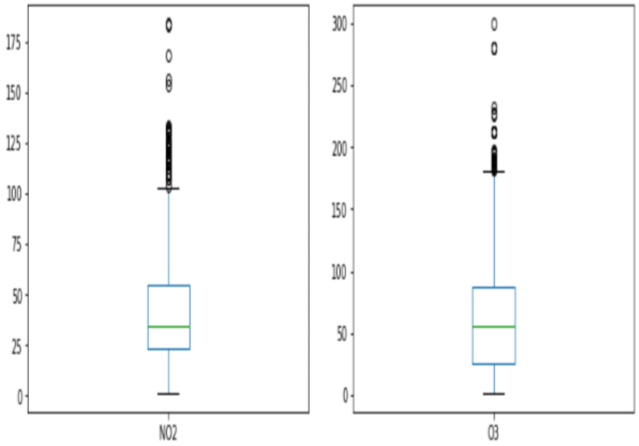


Figure 2: Box Plots for NO_2 and O_3 parameters

III. ARTIFICIAL NEURAL NETWORKS

The prediction of pollutants in the atmosphere is implemented as a regression problem by constructing an ANN(Artificial Neural Network). The gradient descent algorithm which is usually a loss optimization algorithm is achieved using the method of back propagation. The delta rule is used to calculate the error gradient for a feed forward network. For this sole purpose, the delta rule has been implemented.

Delta rule is a gradient descent learning rule for updating the weights of the inputs to artificial neurons in a single-layer neural network. It is a special case of the more general backpropagation algorithm. The back propagation is done in the ANN's having multiple layers. The optimized values that are obtained in a specific set of hidden layers, is selected. In the ANN, the hidden layers act as feature detectors. Completely abiding by the universal approximation theory, a single hidden layer having a large number of neurons can estimate any smooth, measurable function in between the vectors of input and output by choosing a valid set of connecting weights and transfer functions. The Figure 3, represents the structure of Artificial Neural Networks.

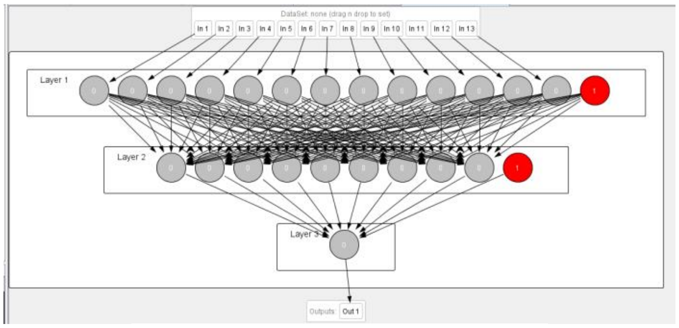


Figure 3: Artificial Neural Networks

The set of continuous functions is introduced in ANN, thus we can implement different number of hidden layers ranging from 9 to 13 for both NO_2 and O_3, but only represented the values for n=10(number of hidden neurons). The RMSE values for different hidden layers is represented in the Table 1. The performance of each of these ANN, is verified and validated by the Root Mean Square Error(RMSE). The RMSE vs epochs of NO_2 and O_3 are represented in Figure 4 and Figure 5 respectively.

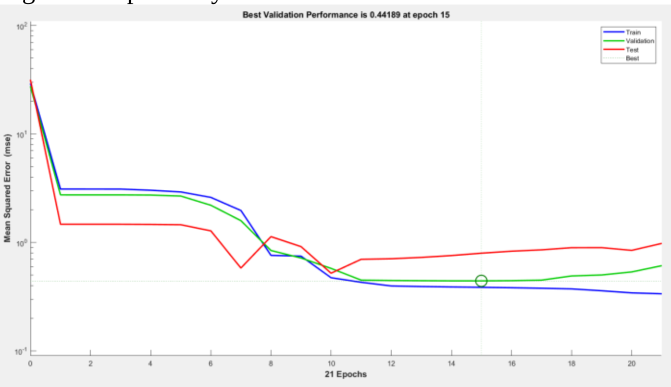


Figure 4: ANN - MSE vs Epochs for NO_2

Best Performance = 0.44189 at Epoch 15

The best validation performance at various epochs is obtained and compared with the ones with different number of hidden layers and these representations are made in the graphs. The Mean Square Error for various hidden layers is considered one after the other and is used for solving a regression problem. When the performance of the ANN drops at a certain number of layers, we consider that it is not the optimal layer count for the given input parameters. At various epochs the values change rapidly and we need to identify the epoch at which we can obtain maximum performance.

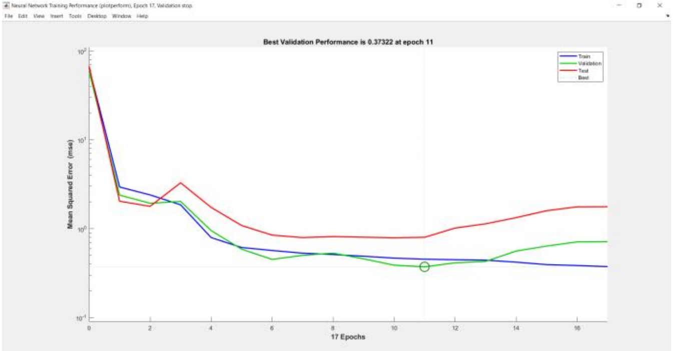


Figure 5: ANN - MSE vs Epochs for O_3

Best Performance = 0.37322 at Epoch 11

The validation results obtained for NO2 and O3 are only considered because, these two are the key components that contribute to air pollution. The Best validation performance is found out by plotting the graph between Mean Square Error and Number of Epochs.

The training and Validation States along with the error histograms are represented in Figure 6 and Figure 7 respectively. These values give us an overall idea of how all the normalized values lead us to the prediction of the levels of the pollutants in the atmosphere for the next 8 hours. The values are indicated using various colors. To reduce the MSE we go for another type of reduction method which is called as PCA-ANN.

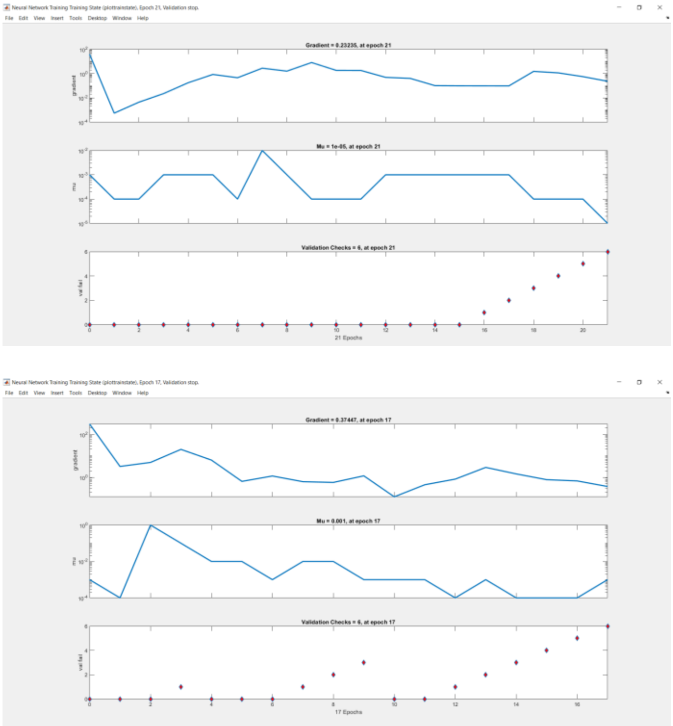


Figure 6: Training and Validation states of NO2 and O3

In order to check the status of the contents of the atmosphere, there is a need to check the error gradients quite often. This improves the prediction parameters of NO2/O3. Hence the error gradient parameters will be very useful.

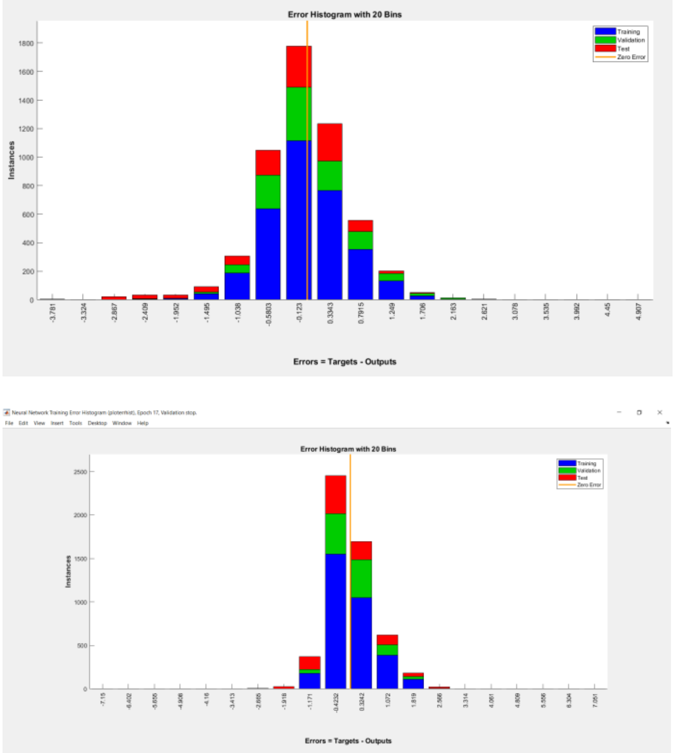


Figure 7: Error Histograms of NO2 and O3

Now, we draw **regression plots** between the targets and outputs to check the type of relationship.

Regression plots are also useful for detecting outliers, unusual observations, and influential cases.

If R=1, there is an exact linear relationship between outputs and targets.

If R is close to 0, then there is no relation between outputs and targets.

The regression plots of NO2 and O3 are as represented in the Figure 8 and Figure-9 respectively.

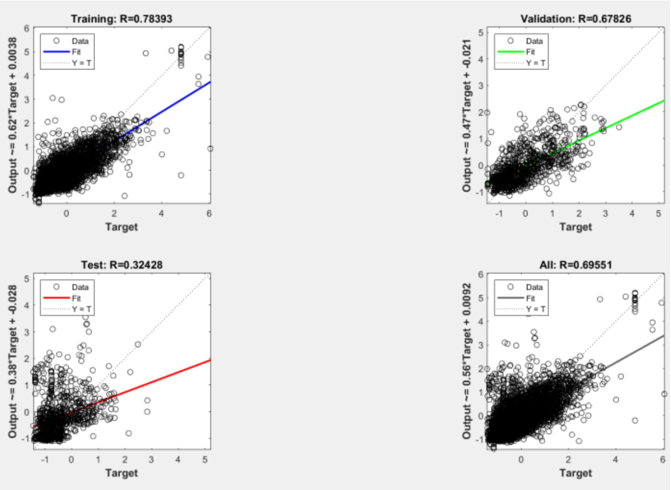


Figure 8: Regression Plots of NO_2

The cluster of the data indicates that the concentrations of the pollutants are confined to a specific range of values and there are some of the points outside the cluster, in indicate that they are not correct readings of the atmospheric contents.

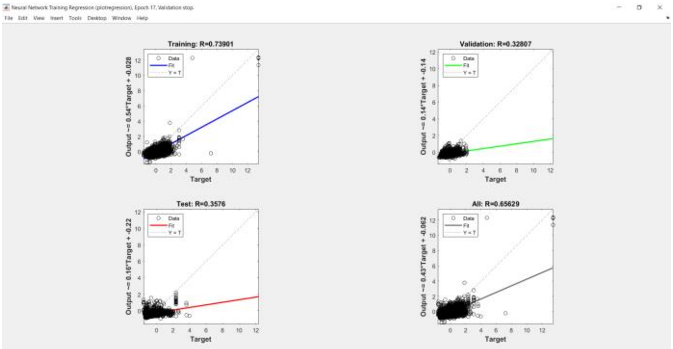
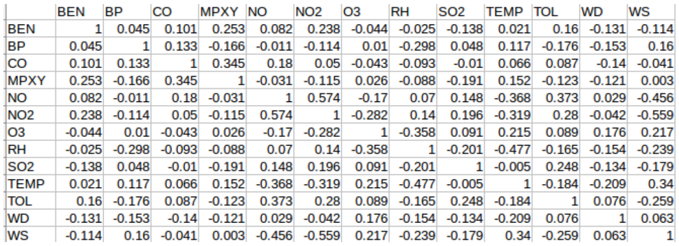


Figure 9: Regression Plots of O_3

PCA-ANN:

The Principal Component Analysis(PCA) is basically used to reduce the number of inputs by providing the schematic view of functional dependencies, variety of data and obtaining information for prediction. The primary role of PCA in the Artificial Neural Network is to minimize the number of predictor variables and transform them into a new set of variables which are defined as *Principal Components*. A correlation matrix is constructed stating all the components and the relation between each other and this matrix computes the Principal Components(PCs).

The PCs corresponding to the largest of eigen values represent the linear combination of variables, which accounts for the maximum total variability in the data. After computing the PCs, the initial dataset is changed/converted/transformed into orthogonal set by multiplying the set o eigen vectors. The PCs whose cumulative amount of variance are approximately 90% are used in the model and remaining components were excluded. Hence only 9 new variables (PCs) were used instead of the original 13 variables. The above reduced set of input variables will be passed as input to the ANN. The mean square error was used to evaluate the prediction result for NO2 and O3 for this model with different number of hidden neurons. The correlation matrix for the 13 input parameters is represented in the Figure 10. The PCA-ANN result for O_3 has the least MSE than original ANN.



	BEN	BP	CO	MPXY	NO	NO2	O3	RH	SO2	TEMP	TOL	WD	WS
BEN	1	0.045	0.101	0.253	0.082	0.238	-0.044	-0.025	-0.138	0.021	0.16	-0.131	-0.114
BP	0.045	1	0.133	-0.166	-0.011	-0.114	0.01	-0.298	0.048	0.117	-0.176	-0.153	0.16
CO	0.101	0.133	1	0.345	0.18	0.05	-0.043	-0.093	-0.01	0.066	0.087	-0.14	-0.041
MPXY	0.253	-0.166	0.345	1	-0.031	-0.115	0.026	-0.088	-0.191	0.152	-0.123	-0.121	0.003
NO	0.082	-0.011	0.18	-0.031	1	0.574	-0.17	0.07	0.148	-0.368	0.373	0.029	-0.456
NO2	0.238	-0.114	0.05	-0.115	0.574	1	-0.282	0.14	0.196	-0.319	0.28	-0.042	-0.559
O3	-0.044	0.01	-0.043	0.026	-0.17	-0.282	1	-0.358	0.091	0.215	0.089	0.176	0.217
RH	-0.025	-0.298	-0.093	-0.088	0.07	0.14	-0.358	1	-0.201	-0.477	-0.165	-0.154	-0.239
SO2	-0.138	0.048	-0.01	-0.191	0.148	0.196	0.091	-0.201	1	-0.005	0.248	-0.134	-0.179
TEMP	0.021	0.117	0.066	0.152	-0.368	-0.319	0.215	-0.477	-0.005	1	-0.184	-0.209	0.34
TOL	0.16	-0.176	0.087	-0.123	0.373	0.28	0.089	-0.165	0.248	-0.184	1	0.076	-0.259
WD	-0.131	-0.153	-0.14	-0.121	0.029	-0.042	0.176	-0.154	-0.134	-0.209	0.076	1	0.063
WS	-0.114	0.16	-0.041	0.003	-0.456	-0.559	0.217	-0.239	-0.179	0.34	-0.259	0.063	1

Figure 10: Correlation Matrix for PCA

Starting from the PCA-ANN we have only put the Epoch values and the corresponding MSE values and the regression plots for both NO_2 and O_3 only For reference to the error histograms, training states please refer to "**project code**", "**working screenshots**" in the submitted set of repositories for further working details. In the PCA-ANN, the MSE vs Epochs determines the performance of this method. The graphs are as drawn in Figure 11 and Figure 12 for NO_2 and O_3 respectively.

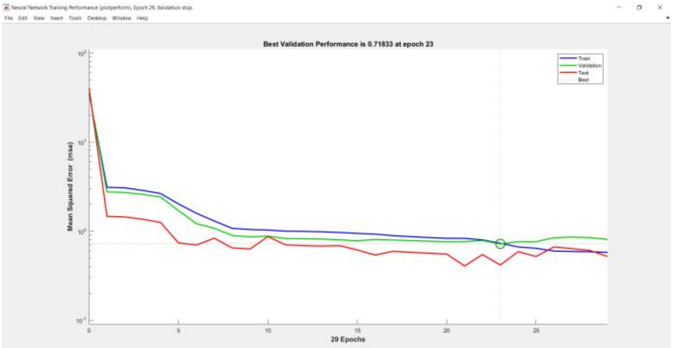


Figure 11: PCA -ANN – MSE vs Epochs for NO_2

Best Performance = 0.71833 at Epoch 23

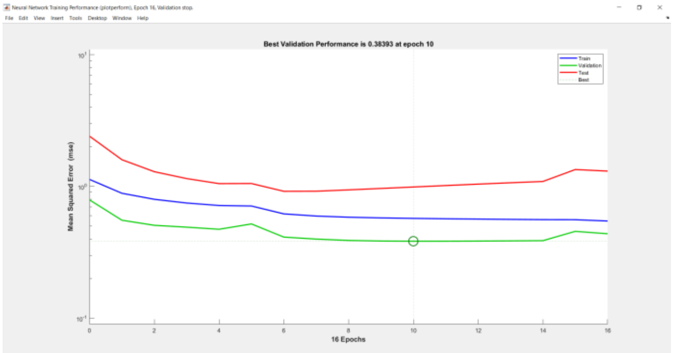


Figure 12: PCA-ANN – MSE vs Epochs for O_3

Best Performance = 0.38393 at Epoch 10

The value at epoch 23 and epoch 10 yields the best performance for all the datasets that are being used for training, testing and also validation. The variations at different epochs are unique for a particular interval of time.

The regression plots of NO_2 and O_3 are represented in the Figure 13 and Figure 14 respectively.

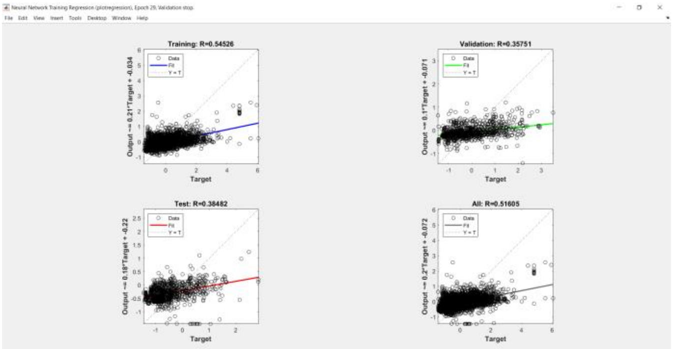


Figure 13: Regression plot for NO_2

The correlation coefficients of each of the parameters of NO_2 and O_3 are modelled as a regression problem. The Mean Square Error corresponding to each of these values and the regressions corresponding to these values are listed as a table. The contents of BEN-BEN, BP-BP, and so on – so forth, WS-WS corresponding to each regression value is 1 for each group. The validation values are obtained from the regression plots and thus give the best performances at each Epoch. The maximum value is 1 and range lasts from -1 to +1. The 13 parameters are represented in the figure of correlation matrix.

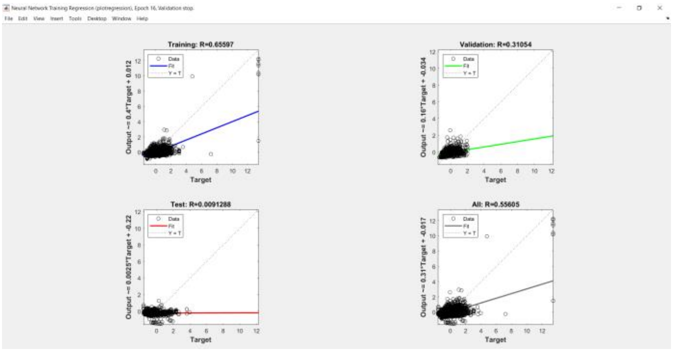

Neural Network Training Regression (plotregression), Epoch 16, Validation stop.
File Edit View Insert Tools Desktop Window Help
Training: R=0.65597
Output ~= 0.4*Target + 0.012
Validation: R=0.31054
Output ~= 0.16*Target + -0.034
Test: R=0.0091288
Output ~= 0.0025*Target + -0.22
All: R=0.55605
Output ~= 0.31*Target + -0.017
Data
Fit
Y = T
Target
0
2
4
6
8
10
12

Figure 14: Regression Plot for O_3

Inference:

Hence we obtained the best performances as follows -

For NO2, at Epoch 15 having 10 hidden neurons has the best performance of 0.44189(ANN)

For O3, at Epoch 11 having 13 hidden neurons has the best performance of 0.36781(PCA_ANN)

IV. ADAPTIVE NEURO FUZZY INFERENCE SYSTEMS

The ANFIS method uses a hybrid architecture composed of a Fuzzy Inference System(FIS) enhanced with the ANN features. The ANFIS uses the key element of human intelligence which is adaptability and learning. Integrating the ANN part into a fuzzy inference system enhances the FIS part with learning/adapting capabilities. The training data determine restrictions on the design methods of the rule base and membership functions.

The FIS has five functional units:

A. Fuzzification Unit(Crisp value to Fuzzy set)
B. The database unit(containing the description of membership functions for input/output variables)
C. A rule base Unit (All the rules defined for FIS)
D. Defuzzification unit (from fuzzy set to a crisp value)
E. The decision unit(performing the inference operations on the fuzzy rules)

The method of subtractive clustering has been used to generate the initial FIS structure, i.e., rules whose antecedent and consequent parameters are then tuned using neural network. This method is used because there are various layers that consists of clusters of same types that lead to the repeated analysis of the same concentration of the pollutants in the atmosphere. The subtractive clustering is employed by a set of innate concentrations which can be scaled to a certain value to perform hybrid learning algorithm(H) and back propagation(B) with gradient and Least Square methods.

The integration of both Neural Networks and Fuzzy Logic principles has tha ability to hold and withstand the advantages of both these in a single frame work. This inference system corresponds to a specific set of Conditional statements also called as IF-THEN rules having the learning capacity to approximate the learning capacity and capability of the non-linear functions in each system design. This can be collaborative and collective at the same time.

The ANFIS layer representation is represented in Figure 15 below.

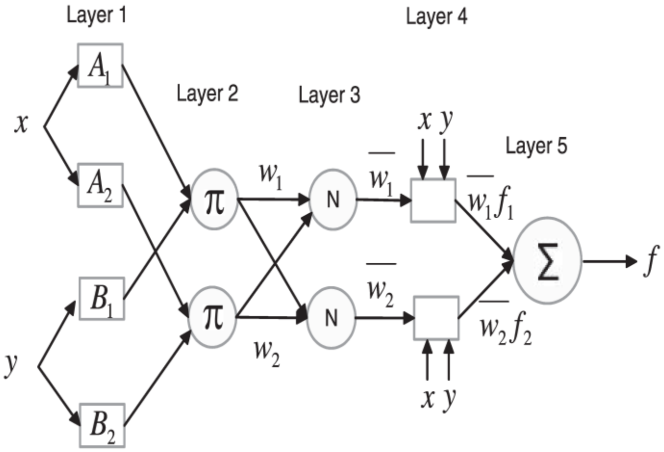


Figure 15: ANFIS Structure

The ANFIS architecture has five layers, with Takagi-Sugeno rules.

1. The first layer (adaptive) forms the premise parameters (the IF part with inputs and their membership functions).
2. The second layer computes a product of the involved membership functions.
3. The third layer normalizes the sum of inputs.
4. In the fourth layer, the adaptive i-node computes the contribution of i-th rule to ANFIS output, forming the consequent parameters (the THEN part with output and its membership function).
5. The fifth layer makes the summation of all inputs.

This layer representation for our system is represented in the Figure 16 below.

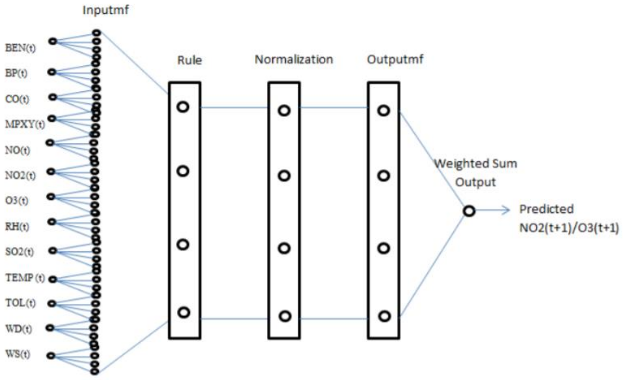


Figure 16: Layer Representation of our system

ANFIS applies a hybrid learning algorithm (H) or backpropagation (BP) algorithm. The hybrid learning algorithm(H) identifies premise parameters with gradient method (min f(x)).

In the ANFIS system, each input parameter might be clustered into several class values to build up fuzzy rules, and each fuzzy rule would be constructed using two or more membership functions. Several methods have been proposed to classify the input data thus making the rules, like grid partition and subtractive fuzzy clustering

The consequent parameters with least square method. At feedforward propagation step from H, the system output reaches layer 4, and the consequent parameters are formed with least square method. With backpropagation (BP) optimization method, the error signal is fed back and the new premise parameters are computed through gradient method. The scatter plots for training and testing of the data are of NO_2 and O_3 are represented in Figure 17 and Figure 18 respectively.

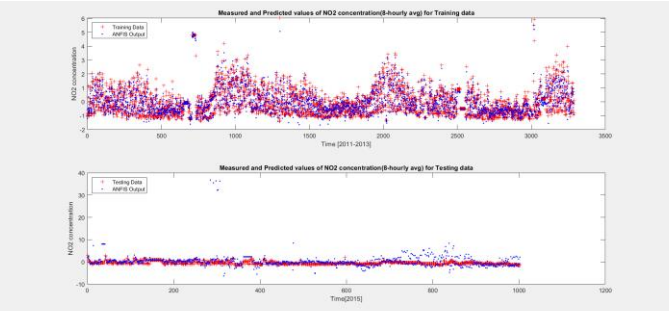


Figure 17: Scatter Plots for Training and Testing(NO_2)

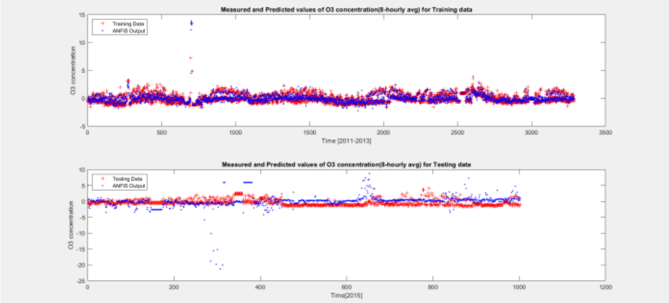


Figure 18: Scatter Plots for Training and Testing(O_3)

. The RMSE Vs Epochs are represented in the graphs NO_2 and O_3 in Figure 19 and Figure 20 respectively.

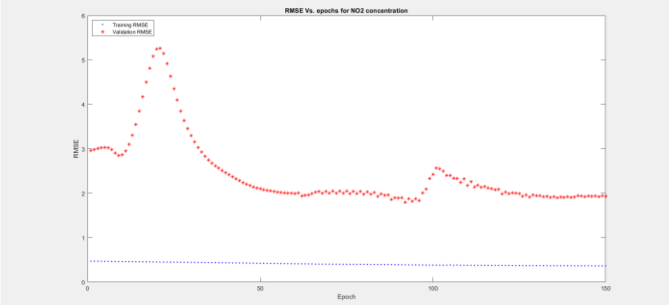


Figure 19: RMSE vs Epoch(NO_2)

PCA-ANFIS: When there are a few input variables, grid partition is a suitable method for data classification. But in this research because of huge amount of input variables, this method cannot be used same as in the case of ANFIS but with Principal Components. For example by having 13 input variables and 4 MFs for each input variable, the rules will be 67,108,864 rules(4^{13}) that create hindrance in the calculation of parameters. Therefore, we have used subtractive fuzzy clustering in order to establish the rule base relationship between the input and output variables according to the number of clusters from the dataset

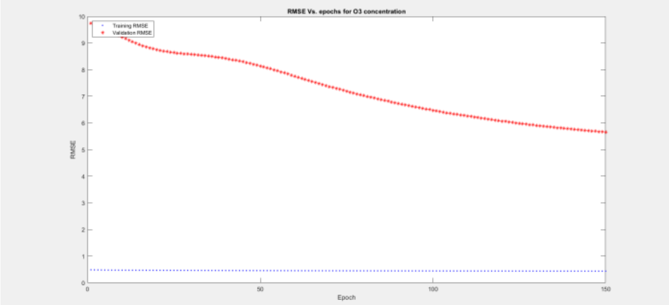
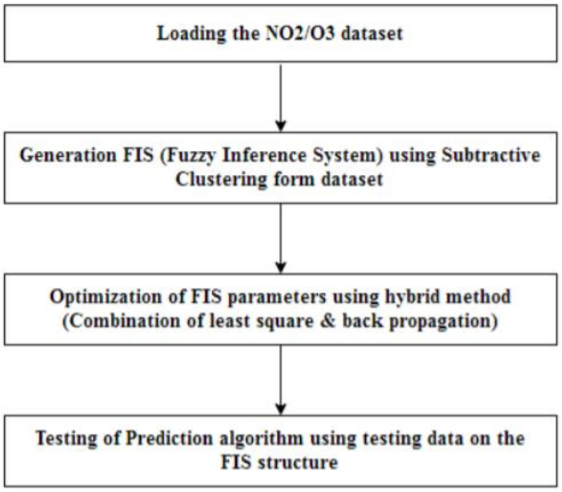


Figure 20: RMSE vs Epoch(NO_2)

The flow of ANFIS/PCA-ANFIS is represented is drawn as a flow chart in Figure 21.



Loading the NO2/O3 dataset
Generation FIS (Fuzzy Inference System) using Subtractive Clustering form dataset
Optimization of FIS parameters using hybrid method (Combination of least square & back propagation)
Testing of Prediction algorithm using testing data on the FIS structure

Figure 21: The flow of ANFIS deployed in our system

The Regression plots of NO_2 and O_3 for ANFIS are represented in the Figure 22 and Figure 23 respectively,

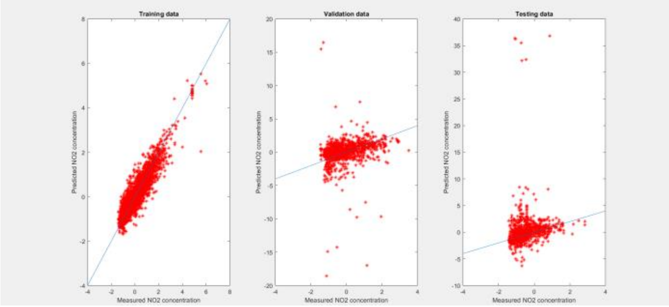


Figure 22: Measured Concentration of NO_2

The training data that is gathered is normalized in the preprocessing phase even for giving the input for ANFIS. When the actual input is given, we give only 9 inputs because the number of inputs is reduced by subtractive clustering which in turn modifies the requirements of the ANFIS. This Feistel structure has all the corresponding weights and activation functions necessary to find out the performance of our system as per the inputs provided in the training, testing and validating data sets.

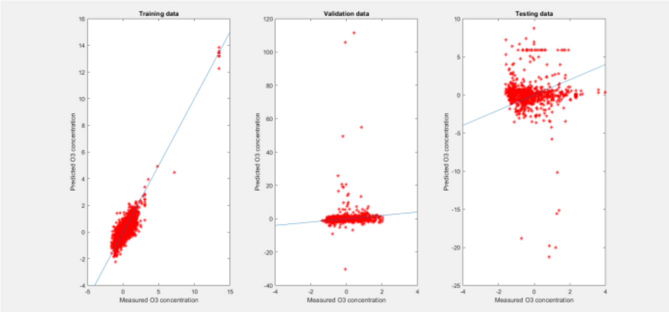


Figure 23: Measured Concentration of O_3

The training data that is gathered is normalized in the preprocessing phase even for giving the input for ANFIS. When the actual input is given, we give only 9 inputs because the number of inputs is reduced by subtractive clustering which in turn modifies the requirements of the ANFIS. This Feistel structure has all the corresponding weights and activation functions necessary to find out the performance of our system as per the inputs provided in the training, testing and validating data sets.

The Scatter Plot representations of the concentrations of NO_2 and O_3 for both training and testing data is represented for PCA_ANFIS as shown in the Figure 24 and Figure 25 respectively.

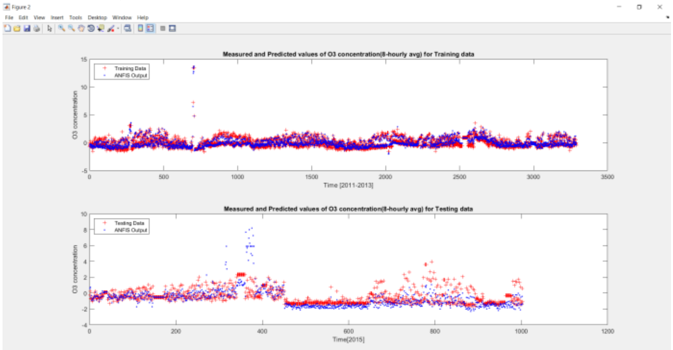


Figure 2
File Edit View Insert Tools Desktop Window Help
Measured and Predicted values of O3 concentration(8-hourly avg) for Training data
Training Data
ANFIS Output
O3 concentration
15
10
5
0
-5
0
500
1000
1500
2000
2500
3000
3500
Time [2011-2013]
Measured and Predicted values of O3 concentration(8-hourly avg) for Testing data
Testing Data
ANFIS Output
O3 concentration
10
8
6
4
2
0
-2
-4
0
200
400
600
800
1000
1200
Time[2015]

Figure 24: Scatter Plots for Training/Testing of PCA_ANFIS(NO_2)

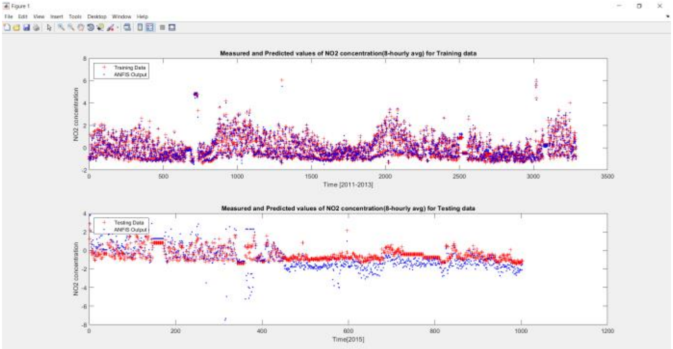


Figure 1
File Edit View Insert Tools Desktop Window Help
Measured and Predicted values of NO2 concentration(8-hourly avg) for Training data
Training Data
ANFIS Output
NO2 concentration
8
6
4
2
0
-2
0
500
1000
1500
2000
2500
3000
3500
Time [2011-2013]
Measured and Predicted values of NO2 concentration(8-hourly avg) for Testing data
Testing Data
ANFIS Output
NO2 concentration
4
2
0
-2
-4
-6
-8
0
200
400
600
800
1000
1200
Time[2015]

Figure 25: Scatter Plots for Training/Testing of PCA_ANFIS(NO_2)

The best validation of the system is measured by the RMSE which is obtained at a particular epoch. The RMSE values are drawn in graphs for NO_2 and O_3 of the PCA_ANFIS is represented in Figure 26 and Figure 27 respectively. This is a unique representation from the existing ANFIS technique.

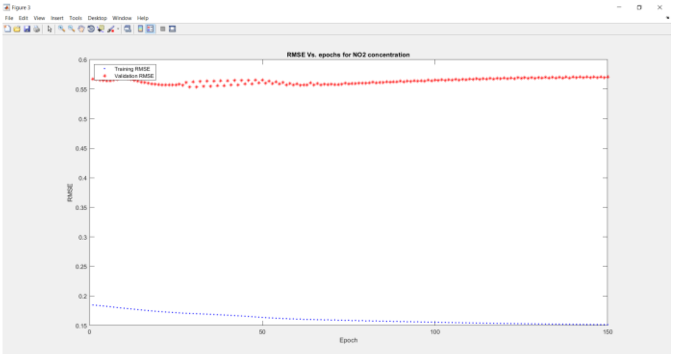


Figure 26: PCA_ANFIS - RMSE vs Epoch for NO_2

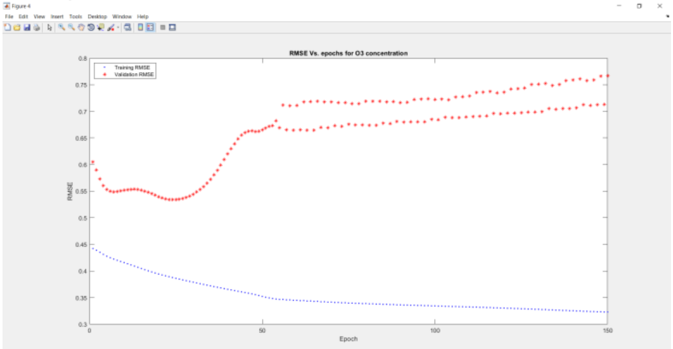


Figure 27: PCA_ANFIS – RMSE vs Epoch for O_3

We construct the regression plots for both NO_2 and O_3 in order to find the relationship between the output and targets. The linear relationship exists only when R=1. We construct different regression plots for training, testing and validation data to find out the outliers in each of these datasets.

The Regression Plots for NO_2 and O_3 for ANFIS is shown in the Figure 28 and Figure 29 respectively and are represented uniquely.

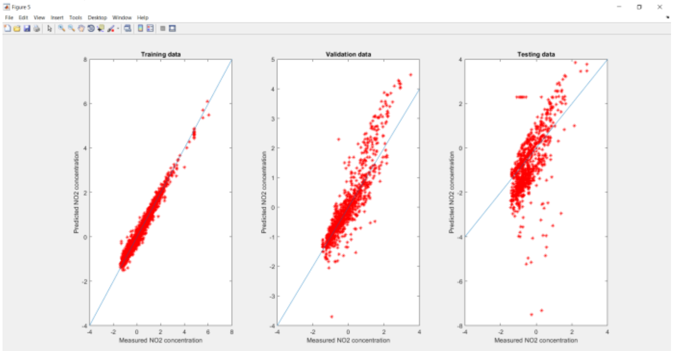


Figure 5
File Edit View Insert Tools Desktop Window Help
Training data
Validation data
Testing data
Predicted NO2 concentration
Measured NO2 concentration

Figure 28: PCA_ANFIS – Regression Plots for NO_2

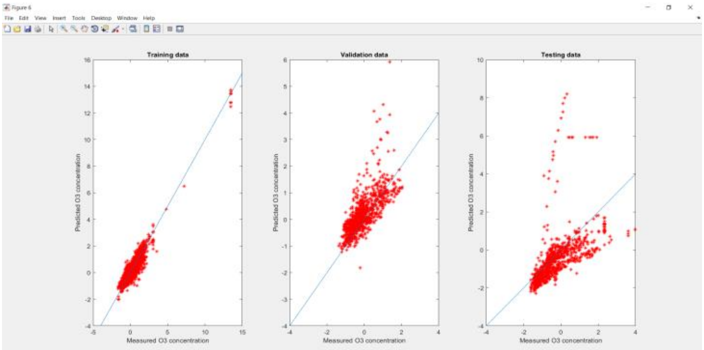


Figure 6
File Edit View Insert Tools Desktop Window Help
Training data
Validation data
Testing data
Predicted O3 concentration
Measured O3 concentration

Figure 29: PCA_ANFIS – Regression Plots for O_3

Inference:

The PCA-ANFIS and ANFIS simulate similar results and the values obtained for RMSE of NO_2 and O_3 are 0.5031 and 0.5215 respectively for the training data. The PCA-ANN is more efficient for testing data

V. CHALLENGES FACED

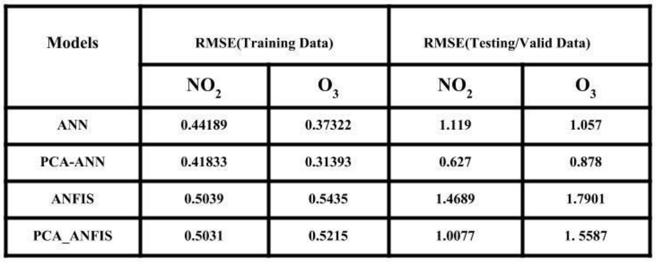
The data collected related to the environment are highly complex to model because of the existing correlation in many variables of unique types which give an intricate mesh of the relationships among them. Apart from this, the factor of knowledge in the database is not consistent, uncertain and at times which is completely inaccurate due to problem with the apparatus and incompetent data collection. The major challenge occurs when the estimation of the high pollutant concentration in metropolitan cities and so the factor of the prediction is inappropriate and this was improved rapidly by us to counter the potential problems.

The higher level in-charges and the buzz among the public to forecast the air quality, to obtain the permissible levels of the atmosphere has led to proper consideration of every possible parameter and robustness to the environmental conditions. The grading of the values made it difficult to normalize because of the variety of data that is being obtained as a series of values picked up by the various sensors. The hourly intervals were hard to collect and implement as well.

The Inherent time series models are generally non-stationary and is the main reason for the occurrence of the non-linear behavior. The dedicated space for the vector intervals is quite large and became difficult to sort out each model at the end.

VI. CONCLUSION

All the models are compared based on the obtained RMSE values. These values are tabulated below –



Models	RMSE(Training Data)		RMSE(Testing/Valid Data)	
	NO_2	O_3	NO_2	O_3
ANN	0.44189	0.37322	1.119	1.057
PCA-ANN	0.41833	0.31393	0.627	0.878
ANFIS	0.5039	0.5435	1.4689	1.7901
PCA_ANFIS	0.5031	0.5215	1.0077	1. 5587

Overall Inferences:

Implementing all the four models, we finally conclude that

1. The PCA-ANFIS is the most efficient model for both NO_2 and O_3 for the training the data.
2. The PCA-ANN is the most efficient model for both NO_2 and O_3 during the implementation of testing and validating data.

Best Model:

Hence on considering all the above obtained values the best method that can be used for the prediction of the contents of NO_2 and O_3 is PCA_ANFIS and in some cases the PCA_ANN is more efficient(but very few times) and the number of hidden neurons that can be used is n = 10. The alerts are generated if the levels are beyond permissible levels of NO_2 and O_3 using PCA_ANFIS.

VII. REFERENCES

1. Ming Cai , Yafeng Yin , Min Xie - Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach in 2008
2. M. Oprea , S. F. Mihalache , M. Popescu – A comparative study of computational intelligence techniques applied to PM2.5 air pollution forecasting in 2016
3. F Grazzini, A Persson, User Guide to ECMWF forecast products, version 4.0 (Meteorological Bulletin M3.2) in 2007
4. J.-S.R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," IEEE Transactions on Systems, Man, and Cybernetics (Volume: 23, Issue: 3, May/Jun 1993)
5. Y.-Q. Zhang, Fu-lai Chung, "A fuzzy neural network tree with heuristic backpropagation learning**,"** Neural Networks, 2002. IJCNN '02. Proceedings of the 2002 International Joint Conference on Volume 1, 12-17 May 2002 Page(s):553 - 558