Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 5, June 2021: 1288– 1298

Research Article

# Particle Swarm Optimisation based Machine Learning Algorithms for the Assessment of Air Quality Level

M. Nelgadevi[1], Dr. Grasha Jacob[2]

## Abstract

Air pollution refers to the release of pollutants into the air that jeopardises human health and also results in serious environmental problems. Many countries are suffering from heavy air pollution. The inhalation of pollutant will lead to acute conditions like asthma, chronic lung diseases, and cancer. The Air Pollution not only harms the health of both humans and animals but also abolish the life of the plants. Hence the estimation of Air Quality is an essential. This paper employs air quality estimation using Particle Swarm Optimization. Multilayer perceptron type of back-propagation Neural Network is used to analyse the status of air pollution at various locations in India. The efficiency of the proposed model is evaluated with K-Nearest Neighbor, Random Forest and Support Vector Machine algorithms by classifying the data into six classes of pollution levels, which can be simply understood by the public as Good, Satisfactory, Moderately Polluted, Poor, Very Poor and Severe.

***Keywords*** – Air Quality Index, Multilayer Perceptron, Particle Swarm Optimization, Random Forest, Support Vector Machine, K-Nearest Neighbor, Pollution, Prediction.

## 1. Introduction

Air pollution is a familiar environmental problem especially connected with the urban areas around the world. In India, the environmental conditions have gradually decayed due to urbanization, development of many factories, lack of awareness, poor maintenance of motor vehicles and damaged road conditions. Air Pollution is estimated to have killed 1.5 million people every year. It is the fifth largest potential killer in India. India has the world's highest death rate from chronic respiratory diseases and asthma, according to the World Health Organization. Heavy vehicular traffic and industrial emissions are the major reasons for the pollution. Hence, air gets polluted with the mixture of gases and has an adverse effect on health which is a major issue worldwide. For many years, improving the quality of air is a challenging task all over the world. So, one should be aware of the status of air inhaled.

[1]Research Scholar, Register Number : 18121172162005
[1]Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 627 012, Tamil Nadu, India
[2] Associate Professor & Head [1, 2] Department of Computer Science Rani Anna Government College for Women ,Tirunelveli 627 008, Tamil Nadu, India.grasha.devi@gmail.com

For this, accurate methods are essential for the assessment of air in advance. Currently Air Quality status has been observed through a large amount of data. It is imperative that particulars on air quality should be given to the public domain in simple terms that is easily understood by a common man. Air Quality Index (AQI) is one such effective tool for the distribution of air quality information to the public.

### 1.1 Air Quality Index

Environmental Protection Agency (EPA) calculates the AQI for main air contaminants such as Particulate Matter, Carbon Monoxide, Ammonia, Sulphur dioxide, Nitrogen dioxide and Ozone. The AQI was grouped, based on the estimation of these pollutants ($PM_{10}$, $PM_{2.5}$, CO, $NH_3$, $SO_2$, $NO_2$, and Ozone). Pollutant concentrations are converted into AQI which assumes values in the range 0–500. The purpose of the AQI is to enable common man to understand how air quality intends to cause harm to health. To make it simpler to comprehend, Table 1 represents partition of AQI into six levels of health concern.

Table 1: AQI with six categories of corresponding colour scheme

| **Air Quality Index (AQI)** | **Associated Health Effects** | **Indicated by this Colours** |
|---|---|---|
| Good<br>(0 to 50) | Minimal impact and no risk | Green |
| Satisfactory<br>(51 to 100) | Minor respiratory discomfort to Sensitive People | Light Green |
| Moderately Polluted<br>(101 to 200) | Affects people with asthma, heart disease, risk for children and elderly people. | Yellow |
| Poor<br>(201-300) | Breathing discomfort to heart patient on prolonged exposure | Orange |
| Very Poor<br>(301 to 400) | Cause respiratory illness to people with lung and heart disease | Red |
| Severe<br>(401 to 500) | Cause respiratory problem even on healthy people, impacts may be experienced by the entire population | Maroon |

A sub-index is calculated for each pollutant, which transforms the raw measurement of individual air pollutants into a single number that may be widely used by the public for air quality communication and decision making whether status of the air is good or bad. The Pollutant with the highest sub-index becomes the Air Quality Index, and is deemed to be responsible pollutant for that day.

## 2. Related Work

In recent years, several types of approaches have been utilised to assess the quality of air for the prediction of air pollution with concern to human health.

Prakash Mamta and Bassin (2010) report the analysis of the ambient air in Delhi city employing air quality index. It has been observed that the calculated AQIs values $SO_2$ and $NO_2$ fall under ‘good’ and ‘good-to-moderate’ categories. The general AQI was found to come under the class 'poor' and 'extremely poor' attributable to RSPM and SPM, separately. The AQI study shows that SPM was in charge of the greatest circumstances in all destinations in Delhi. The dominant part of AQI evaluations of SPM fell under the category of extremely poor [6].

Dragomir (2010) used the K-Nearest Neighbor to categorize the pollution level to forecast the air quality in order to predict the value of the air quality index [4].

Chi-Man Vong et al (2012) constructed the appropriate choice of kernel in Support Vector Machine for forecasting daily ambient air pollutant. The Selection of Linear and Radial Basis Function (RBF) model were relatively good in the prediction of $SO_2$ and $NO_2$. In seasonal test, these models produced superior results with relatively lower errors when compared with other models of SVM [2].

Fontes et al (2013) proposed a Multilayer Perceptron (MLP) with one hidden layer to automate the classification of the impact of air quality on human health, by using only traffic and meteorological data as inputs. The results reveal that an MLP with 40 to 50 hidden neurons and trained with the cross-entropy cost function which can be considered as a good generalization to predict air quality impacts on human health [9].

Xiao Feng et al (2015) presented a novel hybrid model applying air mass trajectory analysis and wavelet transformation into a Multilayer Perceptron (MLP) type of back-propagation neural network to improve forecasting accuracy of daily average concentrations of PM2.5 in two days advance. Combined with meteorological forecasts and respective pollutant predictors, the hybrid model is taken to be an effective tool to increase the predicting accuracy of PM2.5 [10].

Chuanting Zhang, Dongfeng Yuanwe (2015) proposed a fast fine-grained AQI level prediction method based on the implementation of random forest algorithm on Spark. The Method is fast in predicting concentration level of PM2.5 [3].

Saima et al (2017) carried out the Neuro Fuzzy Inference Model to select five air pollutants as inputs and predicted the air quality index on the bases of environment hazard conditions, which impacted on human health as good, moderate, or unhealthy air. The results showed that NF based AQI prediction model classifies the AQI proficiently, strongly, and exactly as compared to conventional method [8].

Akash Saxena and Shalini Shekhawat (2017) applied a mathematical framework to formulate a Cumulative Index (CI) on the basis of an individual concentration of four major pollutants. The Support Vector Machine module Grey Wolf Optimizer has been employed for parameter estimation to classify the quality of air with maximum classification accuracy [1].

Dilbag Singh et al (2017) utilized the Adaptive Neuro-Fuzzy Inference System (ANFIS) and Particle Swarm Optimization (PSO) in the prediction of benzene concentration in the atmosphere. PSO is employed to enhance the accuracy of ANFIS for parameter tuning by calculating multi-objective fitness function which involves accuracy, root mean squared error and correlation [5].

## 3. Proposed Model

With the aim of estimating the quality of air in real-time, an air quality classification model is proposed, based on the four different classifiers of machine learning. The selection of pollutant to differentiate the status of air quality into six classes of pollution level is carried out with Particle Swarm Optimization Algorithm. Air Pollutants data of New Delhi and Chennai of a whole year are chosen to arrive at the main pollutant for the particular place. The Multilayer Perceptron, K-Nearest Neighbor, Random Forest and Support Vector Machine classifiers are compared to evaluate the classification accuracy. The architecture of the system is diagrammatically represented in Fig 1.

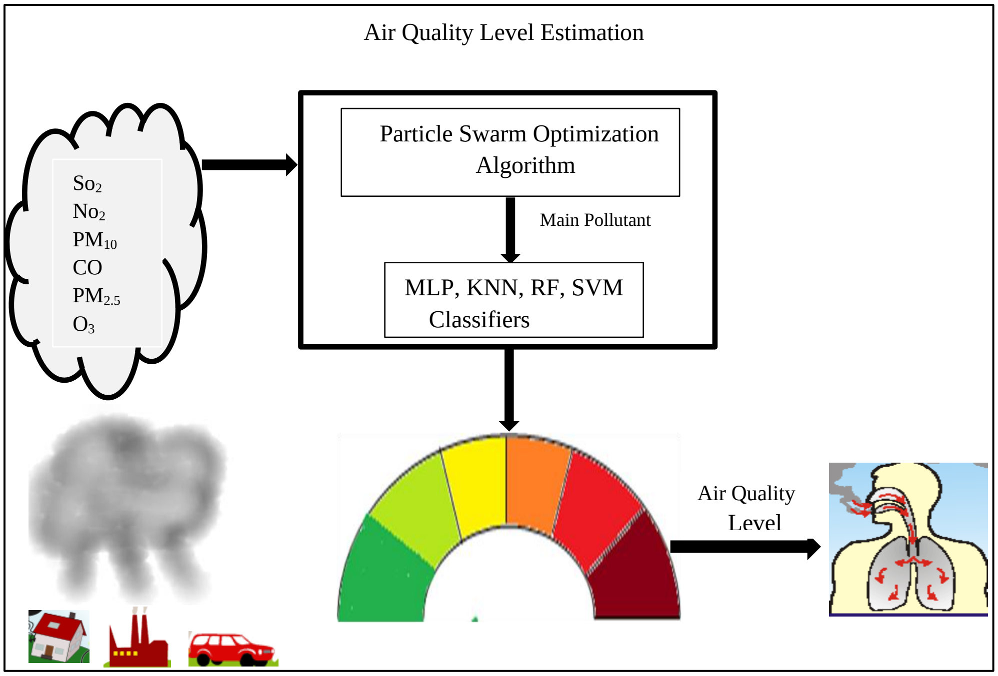


**Fig 1. The Architecture of Air Quality Estimation Model**

### 3.1 Feature Selection using PSO Algorithm

The Particle Swarm Optimization (PSO) algorithm is a swarm based self-adaptive, stochastic optimization technique developed by Dr. James Kennedy and Dr. Russell Eberhart [5]. The algorithm is activated by social attitude of animals such as bird flocking, fish schooling. PSO reiteratively begins by defining the swarm of random solutions. Each potential solution is referred to a particle and set of particles composes a population. The optimization result can be obtained through collaboration and struggling among the individual particles in a population. All particles are initiated randomly with initial velocities to evaluate the objective function at each location of the particle to find the best of each particle and in the entire swarm. The fitness value is calculated for each particle. Each particle keeps track of its position associated with the best fitness in the problem space that has achieved till now, is called pbest. The position associated with the best value obtained so far by any particle among all particles in the population is called gbest. In each iteration, velocity is calculated by using individual and global best positions and each particle positions are updated by the present velocity. Iterations proceed until the algorithm reaches a stopping criterion.

The formula to compute the velocity of particle is shown in Equation (1)

**v = p.v + c1 * r1 (p.best - p.params) + c2 * r2 (gbest.params - p.params) … … … (1)**

The current velocity is computed by adding two elements to the previous velocity of the particle. The first element is the difference between the current position p.params of the particle and the position p.best with the best value obtained by the particle. The second element is computed by the difference between the current position p.params of the particle and the position gbest.params of the best value of all particles of the swarm. The components are multiplied by the learning constants c1 and c2 which remains the same, and greater than one. Each factor is also multiplied by random numbers r1 and r2, respectively. The control parameters used by the PSO algorithm is given in Table 2.The formal systematic flow of major pollutant selection is given in the form of Algorithm Particle Swarm Optimization.

Using this computed velocity, the new position of the particle is assigned in Equation (2)

$$\mathbf{p.params = p.params + v} \qquad \ldots\ldots\ldots \text{(2)}$$

Where

p.v = Velocity of particle
p.best = Individual best position of a particle
gbest.params = Global best position at iteration
p.params = position of a particle
c1, c2 = accelerating constants
r1, r2 = random numbers between 0 and 1

The popularity of the PSO algorithm is its simplicity as it consists of the above two equations only to update the position of the particles.

Table 2: Control parameters of the PSO for feature selection

| Parameters | New Delhi | Chennai |
|---|---|---|
| Population size | 100 | 100 |
| Dimensions | 6 | 4 |
| Error Criterion | 1e-05 | 1e-05 |
| c1 | 2 | 2 |
| c2 | 2 | 2 |
| Iterations | 100 | 100 |

**The Main pollutant Extraction using Particle Swarm Optimization**

**Input: List of Pollutants (**$PM_{10}$, $PM_{2.5}$, CO, $NH_3$, $SO_2$, $NO_2$, and Ozone**)**
**Output: Main Pollutants**

```
Generate initial population using random selection of pollutants (pop)
For each particle of population
    Initialize the velocity and position
End For
Do
  For each pop
    Compute the fitness value using the classifier MLP,K-NN, RF, SVM
     If obtained fitness value is effective than pBest then
       pBest = obtained fitness value
     End If
  End For
  Select the particle as gBest in best of PBest
  For each pop
    Compute velocity by equation (1)
    Move position by equation (2)
  End For  While Termination Condition not met
```

### 3.2 Classification of Air Quality Level

### a) Multilayer Perceptron

A Multilayer Perceptron (MLP) type of back-propagation neural network is applied for the classification of pollution into different levels. MLP comprises layered processing elements called neurons. It contains one input layer, one or more hidden layers and one output layer. Every neuron is connected to all the other neurons of hidden layer or the output layer. Among the collected data, training sets are used to adapt the weight to train the model which utilises the back propagation as learning technique. The enough number of neurons in the hidden layers helps reach the target. The testing sets are used to predict the performance based on the trained model. The Response of the MLP is demonstrated by the output layer. An MLP with three hidden layers can be represented graphically in Fig 2. In this work, the neuron in the input layer corresponds to the number of pollutant concentrations that feed as inputs. Three hidden layers are constructed, each with thirteen hidden neurons. The output is computed at each node in middle of the network. The weights are adjusted in each layer to predict the correct pollution level in the output layer. The value at the hidden layer is calculated using equation 3. The equation 4 is to pass the output to the next layer.

$$\text{Value at h1} = xi * wi + \ldots + xn * wn + bi \qquad \ldots\ldots\ldots (3)$$

Where,

x1 …..xn - input to the neurons
w1…. wn - the correspond weights
b - bias value

$$\text{Activation function} = \frac{1}{1+e^{-h1}} \qquad \ldots\ldots\ldots (4)$$

The value at each layer is evaluated and the output obtained in output layer is compared with the target value to calculate the error and is done using the Equation 5.

$$\text{Error} = \frac{1}{2}(\text{target} - \text{output})^2 \qquad \ldots\ldots\ldots (5)$$

The Error is back propagated to update the weight and the learning rate is fixed as 0.01 to train the model until the output of the model is close to the target or error is minimized. The learning rate controls to adjust the weight and must be low to get the best result. The pollutants concentrations are given as input in the input layer to test the model to classify the pollution level.

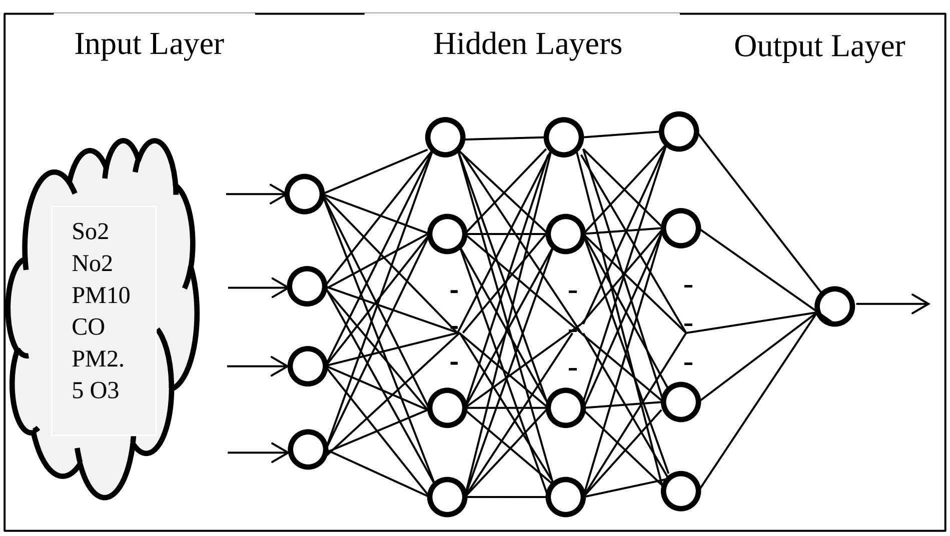


**Fig.2 Multilayer Perceptron with (13, 13, 13) Hidden Layers**

### b) K-Nearest Neighbor

K-Nearest Neighbors (K-NN) is used for Classification of Multi-class of pollution level. K-NN is based on the understanding that it classifies Pollution level for new data based on attributes of sample training dataset, by a majority group of K-nearest neighbours [4]. Hence it does not really achieve a model from training dataset but store the data. Given a new data, its pollution level is obtained by k nearest neighbours of the stored training data set. K=3, pays highest accuracy for this model.

The Euclidean distance is used as the distance metric to calculate the k nearest neighbours and is calculated as.

$$Euclidean\ Distance = \sqrt{\sum_{i=1}^{n} \ (xi - yi)} \qquad \ldots.\ \ldots.\ \ldots.\ (1)$$

The nearest neighbours are the training data that are close to the test data that helps to predict the pollution level for the new dat. Based on the majority vote of the K nearest neighbours, the test data is classified to a particular class of pollution level.

### c) Random Forest

The Random Forest is a supervised ensemble classification Algorithm which uses many decision tree models. The data is partitioned into subsets and each subset leads to a decision tree. The Final outcome is obtained through analysing the majority voting by multiple decision trees. In this study, the training dataset of the pollutants are divided into bootstrap subsets to build independent decision trees. The bootstrap samples are produced by sampling the dataset with replacement, some data could be selected more than once. To build a decision tree the selection of optimal splitting attribute is required in order to get a good classification result for the data subset. The criterion gini index is an attribute selection measure used to select minimum impurity attribute for splitting in the generation of sub tree [3]. The internal nodes of the tree representing the pollutants are categorised until the leaf nodes which represent the class label. The decision obtained by all the trees is compiled to produce the final result. Hence the random forest classifier is trained enough to test with testing dataset. Each decision tree predicts different class labels. The majority voting is considered to estimate the class label for particular pollutant concentrations.

### d) Support Vector Machine

Support Vector Machine (SVM), is a supervised machine learning algorithm that is employed for analysing the dataset for classification purpose. In this study, the input vectors in the training dataset are mapped to high dimensional feature space to obtain the nonlinear class boundaries to perform a multi-class classification. The model is trained with the value of each pollutant being the value of a particular coordinate. The hyper-plane is a line that linearly separates the dataset with corresponding class labels. The Selection of hyper-plane in mapping data to high dimension is represented as kernel [2]. Radial Basis Function is employed in SVM Model to separate the classes as wide as possible. The decisive function of the SVM classifier depends on the selection of support vectors, the data points nearest to the hyper-plane. The regularization parameter must be small to maximize the margin between data points in each class to train the model to classify the training sets correctly. The testing data set is mapped into the same space to calculate the class label for the particular pollutants.

## 4. Experimental Analysis

### 4.1 Study area

The efficiency of the classifier is applied on real data of two major Indian cities, New Delhi and Chennai. Air pollution is frequently degenerating in New Delhi, the capital of India and it is also life-threatening in Chennai, the capital of Tamil Nadu. In New Delhi, poor quality of air damages the lungs of 50 percentages of the children. A recent survey reports that Chennai residents are affected a lot by air pollution and the knowledge of health hazards of pollution is also poor. In one day, a person inhales up to 10,000 crores of suspended particulate matter (SPM) mostly from vehicular exhaust, industrial pollutants, construction material and waste and these particles take less than 30 seconds to enter into bloodstream. This study has collected data from four major places in New Delhi- RK Puram, Mandir Marg, Shadipur and Ihbas Dilshad garden. In Chennai – Alandur, Manali and Velachery are considered for data collection.

Hence this study, concentrates on estimating the status of air quality data with respect to its effect on human health. One year of data is collected from various cities of India for the classification. The dataset contains attributes such as sampling date, state, City, location of monitoring station, and the concentrations of air pollutant. The following criteria pollutants namely CO, $NO_2$, $PM_{2.5}$, $SO_2$ and $O_3$ in Chennai and CO, $NH_3$, $NO_2$, $O_3$, $PM_{10}$, $PM_{2.5}$ and $SO_2$ in New Delhi have been considered for determination of the status of Air Quality. The data set is divided into two sets: 70% for training and 30% of data is used for testing. The Automatic monitoring environmental data have been selected as an input which includes number of observations. This in turn is recorded daily wise 24-hours average concentration value (8-hourly in case of CO and $O_3$) every month during the period of one year in 2017. The data are downloaded from the Central Pollution Control Board (CPCB), Ministry of Environment, Forests & Climate Change, and Government of India. Website: **www.cpcb.nic.in,** which has been established to determine the present air quality status to control and regulate pollution.

### 4.2 Evaluation Metrics

This section discusses the important measures for appraising this air quality estimation Model. TP is the correctly classified pollution level for each class. TN is the correctly unclassified pollution level for each class. FP is the incorrectly categorised pollution level for each class. FN is the incorrectly uncategorised pollution level for each class.

Accuracy is the ratio of correctly classified pollution levels by the classifier to the total number of input samples and is calculated using the Equation 6.

$$\text{Accuracy} = \frac{TP+TN}{Total\ data\ set} \qquad \ldots\ldots\ldots (6)$$

The accuracy obtained for New Delhi and Chennai in the assessment of air quality is shown in Table 3. It shows that Multilayer Perceptron gives 6.5% better accuracy with K-Nearest Neighbor, 7.2% better accuracy than Random forest and 24.5% better when compared to Support Vector Machine. Its Comparative analysis in the performance of the classifier is represented in Figure 3.

Precision is the ratio of correctly classified pollution level to the total classified pollution level by the classifier for that class and is obtained using the Equation 7.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \ldots\ldots\ldots (7)$$

Recall is the ratio of correctly classified pollution level by the classifier to the all observations in that class and is acquired by the Equation 8.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \dots\dots\dots (8)$$

F1Score is computed as the harmonic mean of Precision and Recall. It is calculated with the Equation 9.

$$\text{F1Score}=2 \times \frac{Precision \times Recall}{Precision+Recall} \qquad \dots\dots\dots (9)$$

The FP and FN must be minimized to get better performance of the classifier.

The pollutants which make a clear intra class differentiation are selected using Particle Swarm Optimization. Four classifiers namely Multilayer Perceptron, K-Nearest Neighbor, Random Forest and Support Vector Machine are used for the classification of air quality using the main pollutants. The values obtained for the various metrics using the four Classifiers for New Delhi and Chennai data are tabulated in Table 4 and Table 5. From the result, for New Delhi primary pollutants which have a toll on health is found as Carbon monoxide and Ozone. In Chennai the pollutant which discriminate the status of pollution are Sulphur dioxide and Nitrogen dioxide. It is found that MLP gives the best accuracy in the classification of air pollution into different levels compared to the others.

**Table 3: Classification efficiency of different classifiers in estimating air quality**

| City | Multilayer Perceptron | K-Nearest Neighbor | Random Forest | Support Vector Machine |
|---|---|---|---|---|
| **New Delhi** | **95.19** | 88.00 | 87.20 | 48.80 |
| **Chennai** | **92.30** | 86.53 | 85.89 | 89.74 |

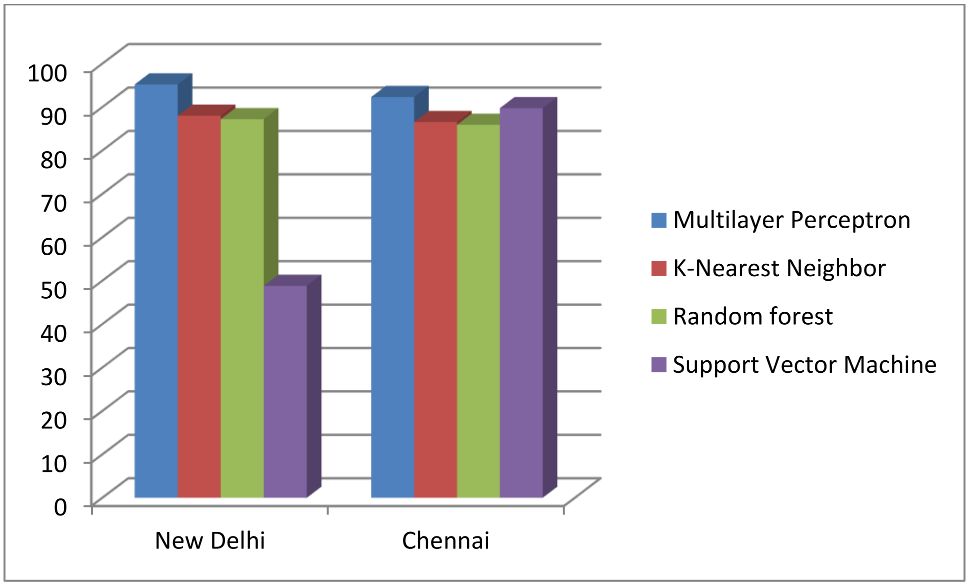


**Fig 3. Comparative analysis of the Classifiers**

**Table 4: Evaluation Measures of the model using four Classifiers for New Delhi dataset**

| Evaluation Measures | Multilayer Perceptron | K-Nearest Neighbor | Random Forest | Support Vector Machine |
|---|---|---|---|---|
| Precision | **94.32** | 86.31 | 76.19 | 46.79 |
| Recall | **95.86** | 87.31 | 67.76 | 28.17 |
| F1score | **94.88** | 85.90 | 66.32 | 25.00 |
| Pollutants | **[CO, $O_3$]** | [CO,NH3] | [NH3,$NO_2$,] | [CO,$NO_2$] |

**Table 5: Evaluation Measures of the model using four Classifiers for Chennai dataset**

| Evaluation Measures | Multilayer Perceptron | K-Nearest Neighbor | Random Forest | Support Vector Machine |
|---|---|---|---|---|
| Precision | **90.85** | 73.84 | 74.67 | 77.02 |
| Recall | **91.64** | 70.92 | 68.25 | 75.85 |
| F1score | **91.16** | 70.93 | 70.29 | 75.95 |
| Pollutants | **[$SO_2$, $NO_2$]** | [CO,$PM_{2.5}$] | [$SO_2$,$NO_2$] | [CO,$PM_{2.5}$] |

## 5. Application

The Government of India has taken several steps to reduce air pollution. The AQI denotes the Quality of air by very simple representation of colour so that a layman can understand. The AQI is an index for informing the daily air quality. It conveys how clean or polluted the air is, and what associated health effects might be an apprehension for the people. The AQI focuses on health effects an individual may experience within a few hours or days after breathing polluted air. Two or more pollutants may be responsible for the pollution in a day and so everyone must be aware of those pollutants to which they are sensitive. This work helps the local residents of two major cities in India – New Delhi and Chennai to plan their outdoor activities when the air quality is better.

## 6. Conclusion

It is inferred that air pollution in India rises every day, results in poor ambient air quality and has a great impact on millions of citizens and the environment. This work presents a review on the air pollutant data to estimate the quality of air which can be easily understood by the public as Good, Satisfactory, Moderately Polluted, Poor, Very Poor and Severe. The results reveal that Multilayer Perceptron based estimation, classifies the air quality significantly better than K-Nearest Neighbor, Random Forest and Support Vector Machine. Particle Swarm Optimization has been employed for pollutant selection with the aim of improving classification accuracy. The model works to the optimum level and hence can be applied to estimate the status of air we inhale in the other cities also.

## REFERENCES

1. Akash Saxena and Shalini Shekhawat.,2017. Ambient Air Quality Classification by Grey Wolf Optimizer Based Support Vector Machine. Journal of Environmental and Public Health. ID 3131083.

2. Chi-Man Vong, Weng-Fai Ip,Pak-kinWong and Jing-yi Yang1.,2012, Short-Term Prediction of Air Pollution in Macau Using Support Vector Machines, Journal of Control Science and Engineering.Volume Article ID 518032, 11 pages

3. Chuanting Zhang, Dongfeng Yuan.,2015. FastFine- Grained Air Quality Index Level Prediction Using Random Forest Algorithm on Cluster Computing of Spark. 978-1-4673-7211-4/15 IEEE.

4. E. G. Dragomir, 2010. "Air Quality Index Prediction using K-Nearest Neighbor Technique", Buletinul Universităţii Petrol – Gaze din Ploieşti, Vol. LXII No.1/ 103 – 108.

5. Husanbir Singh Pannu , Dilbag Singh, Avleen Kaur Malhi ,2017Multi- objective particle swarm optimization-based adaptive neuro-fuzzy inference system for benzene monitoring. Neural Computing and Applications. DOI 10.1007/s00521-017-3181-7. Springer

6. Indresh Kumar Gupta., 2015. A Review on Particle Swarm Optimization. International Journal of Advanced Research in Computer Science and Software Engineering. Volume 5, Issue 4.

7. Prakash Mamta, Bassin ., 2010.Analysis of ambient air quality using Air Quality Index- A Case Study International Journal of Advanced Engineering Technology. E-ISSN 0976-3945.

8. Saima Munawar, Muhammad Hamid, Muhammad Saleem Khan, Ashfaq Ahmed and Noreen Hameed.,2017.Health Monitoring Considering Air Quality Index Prediction Using Neuro Fuzzy Inference Model: A Case Study of Lahore, Pakistan. Journal of Basic & Applied Sciences, 13, 123-132.

9. Tania Fontes, Luis M. Silva, Sergio R. Pereira, and Margarida C. Coelho., 2013.Application of Artificial Neural Networks to Predict the Impact of Traffic Emissions on Human Health**.** EPIA 2013.

10. Xiao Feng, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, Jingjie Wang., 2015.Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment. 107, 118e128