International Journal of Engineering & Technology, 7 (2.9) (2018) 12-17

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Improving air quality management using gradient boosting based hierarchical temporal memory neural networks and fuzzy based classification based regression tree

S. Sagayaraj [1] *, Dr. N. Vetrivelan [2]

[1] Assistant Professor, Department of computer science, Govt. Arts and Science College, Veppanthattai
[2] Professor,Department of computer science, Srinivasan College of Arts and Science,Perambalur
**Corresponding author E-mail: sagayaseelan140@gmail.com*

Abstract

In recent years, air pollution introduces different biological molecules, particulate and several harmful materials which affect the human health and activities. So, the quality of the air should be maintained for avoiding the above issues. To manage the air quality initially the meteorological data have been collected from Ariyalur that includes the condition of air, data collected date, high and low temperature, wind speed, wind direction and relative humidity. The collected data has to be preprocessed by applying the normalization and data mining techniques and those preprocessed data's are used to predict the pollutants and the concentration level of the pollutants such as sulfur dioxide (SO2), carbon monoxide (CO), nitrogen dioxide (NO2), and nitric oxide (NO). Then the particulate matter level in the air has to be predicted by Gradient Boosting based Hierarchical Temporal Memory Neural Networks (BHTMNN). From the predicted value the strength of the pollutants is classified by using the Fuzzy based Classification based Regression Tree (FCART) which is used to recognize the disease arises in the human respiratory system. Then the performance of the proposed system is evaluated using the mean square error, classification accuracy, sensitivity and specificity.

Keywords: *Air Quality; Preprocessing; Air Quality Prediction; Carbon Monoxide (CO); Nitrogen Dioxide (NO2); and Nitric Oxide (NO); Gradient Boosting Based Hierarchical Temporal Memory Neural Networks; Fuzzy Based Classification Based Regression Tree.*

1. Introduction

In the recent years, the growth of the industrial revolution causes several environmental problems. More than 90% environment problem created by the Urbanization process which lead to create the several issues in the human beings [1]. The main issue affected by human beings is air pollution because air contains many substances which may be created by man made or natural process. The air substances introduce most biological molecules, particulars and dangerous material into the atmosphere. The major pollutant of the air is nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO), nitric oxide (NO) and ground level ozone [2]. These dangerous particulars affect living organisms such as human beings, plants and animals and so on. Mostly the people, older adults, children are affected by the these particle pollution exposure and which creates several health issues like, coughing, chest tightness, irritation of the eyes, nose, reduced lung function, asthma attacks and heart attacks and so on. In addition the air pollutants increased the UV radiation, global warming, acid raining and greenhouse effects. To minimize the impact of the air pollutants, the air quality should be predicted [3].

The air quality is predicted by using the several forecasting techniques such as, K-Nearest Neighbor, Classification and Regression Tree (CART), Hidden Markov Model (HMM), Support Vector Machine (SVM), Particle Swarm Optimization Algorithm (PSO), Genetic Algorithm (GA) and Artificial Neural Networks (ANN) [4]. These forecasting approaches reduce the impacts of the air pollutants in the present environment, but it requires the largest amount of meteorological data. The air prediction system requires continuous meteorological data otherwise the small changes affected whole system this one of the major difficulties in the prediction system. So, in the proposed system uses the hybrid soft computing techniques [5]are used to predict the air quality with an efficient manner. To predict the quality of the air, the meteorological data have been collected from Ariyalur which includes the condition of the air, type, data, temperature, wind speed, direction and relative humidity. The collected data have some of the noise which reduces the performance of the predicting system, so it has to be preprocessed by using the normalization which is used to estimate the level of concentration of the biological particles.

Then the particle concentration has been predicted by using the Gradient Boosting based Hierarchical Temporal Memory Neural Networks (BHTMNN) and the strength of the pollutants is classified by the Fuzzy based Classification based Regression Tree (FCART). So, the proposed system overcomes the existing system issues because the neural networks work based on the boosting algorithm which used to improve the learning and adoption rate while predicting the quality of the air. From the predicted air pollutants the disease arises in the human respiratory system is identified. The rest of the section organized as follows, section 2 discusses the several discussion about the air quality managing system, section 3 describes the proposed methodology and section 4 discuss that the results and discussion.

2. Related works

Air pollution is one of the environmental challenges in the developing technology. So, the air pollutants must be controlled by using the air quality prediction system. Zhongliang Yue et al., [6] author discusses that the air quality measure in the time of 2010 Asian Games which is conducted in Guangzhou. The air quality is continuously monitored in November 12th to 21th. During the quality prediction system the different air pollutants such as SO_2,NO_2 and PM_{10} are predicted and the concentration of these pollutants is increased when compared to the 2006,2007, 2008 year air quality measurement. So, the quality of the air must be predicted and controlled the pollutant concentration level. Tian et al., [7] uses the hybrid genetic algorithm and support vector regression tree and hybrid genetic algorithm back propagation neural network for recognizing the air quality. The system monitored the quality of the air in China, Chongqing university electronic and communication engineering department. Then the performance of the quality system is evaluated with the help of the absolute and relative error prediction measure which predict the benzene, CO, NO_2, formaldehyde and toluene in the department. Thus the proposed system summarizes that the air quality should be monitored in the indoor and in-car system.

Ana Russo et al., [8] author uses the stochastic variables to train the artificial neural networks for predicting the air forecasts. The stochastic variable reduces the training time and provide the efficient result when matching the air pollutants in the air forecast. The performance of the system is evaluated with the help of the temporal correlation between the input variables and the stochastic variable which continuously monitored the forecast system and optimize the input while predicting the air pollutants. Suarez Sanchez et al., [9] developing the regression based support vector machine model for predicting the air quality in the Gijon Urban Area. The developed air quality system monitoring the ozone, sulfur dioxide, carbon monoxide, nitrogen oxides and dust in the year from 2006 to 2008. From the measured air pollutants the dependency between the primary and secondary pollutants are monitored and controlled the affected on the human health. Then the performance of the proposed system is evaluated with the help of the correlation coefficient, mean and relative error values.

Maruf Hossain et al., [10] proposed the Hidden Markov model and fuzzy based approach for analyzing the air quality metrics in the hourly based manner. The proposed approach monitors the air substance in continuously which is used to avoid the effects on the human health. Thus the performance of the proposed system is analyzed with the artificial neural networks and the normal fuzzy concepts. Xiao Feng et al., [11] analyzing the quality of the air in 13 different air quality station using the wind direction and speed parameter. The system uses the neural networks to monitor the daily $PM_{2.5}$ concentration level in the air pollution monitoring station like Beijing, Hebe and Tianjin. The wind forecasted parameters are processed by using the multilayer perceptron based back propagation neural networks, which produce the best result when matching the pollutants. Thus the performance of the proposed system is evaluated with the help of the root mean square error value. Thus the proposed system evaluates the concentration of the air pollutants using the Gradient Boosting based Neural Networks and the strength of the pollutant is identified by Fuzzy based CART method. Thus the classified result is used to estimate the level of concentration of the air pollutants.

3. Proposed system

The present technology lead to create environmental problems and health issues to the human beings. One of the main environmental issues is air pollution because it consists of lots of air pollutants which causes of disease to the human beings. So, the quality of the air should be monitored and controlled by using the air quality prediction system which is done with the help of the Gradient Boosting based Hierarchical Temporal Memory Neural Networks (BHT-MNN). The following figure 1 depicted that the proposed overall system architecture.

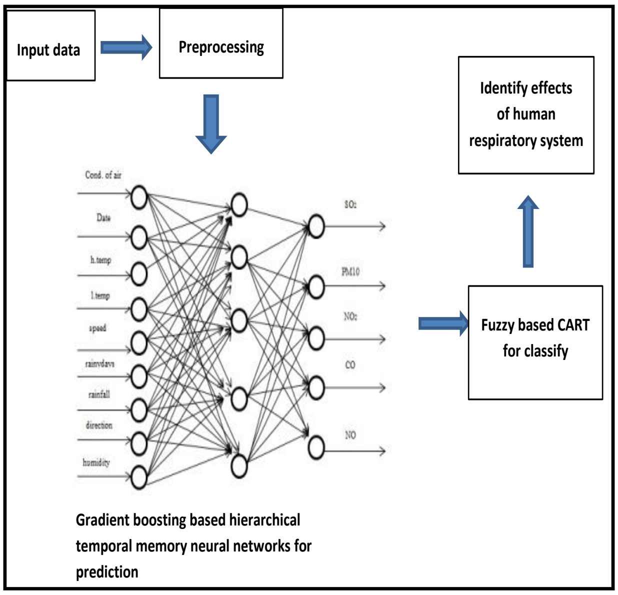
Input data
Preprocessing
Identify effects of human respiratory system
Cond. of air
Date
h.temp
l.temp
speed
rainydays
rainfall
direction
humidity
SO2
PM10
NO2
CO
NO
Fuzzy based CART for classify
Gradient boosting based hierarchical temporal memory neural networks for prediction

Fig. 1: Proposed System Architecture.

The above figure 1 shows that the proposed system architecture which uses the Ariyalur meteorological data's such as wind energy, direction, quality, type, date, temperature and so on. The collected meteorological data set, the quality of the air has been predicted by using the Gradient Boosting based Hierarchical Temporal memory Neural Networks. This neural network classifies the meteorological data with the help of the boosting classifier which promotes the week classifier into the stronger classifier. From the predicted air pollutants, the concentration of the air pollutants is identified by using the Fuzzy based CART classifier. Then the effects on the human respiratory system is evaluated with the help of the resultant pollutants strength.

3.1. Data pre-processing

Data preprocessing is the first stage in the air quality prediction system and it is an important step in the pattern recognition process. In the proposed system Ariyalur meteorological data have been used to estimate the concentration level of the pollutants. The data set consists of following details such as wind direction, wind speed, date, time, temperature details, rain fall details. The gathered meteorological data have some of the noise which lead to reduces the performance of the system. So, the noise has been removed by using the data mining techniques like, missing value replacement and normalization process. In this paper the missing meteorological data are replaced by applying the incremental mean value process [12]which is performed by using following equation 1.

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \mu_{n-1} + \frac{1}{n}(X_n - \mu_{n-1}) \qquad (1)$$

Where, μ_n is the incremental mean value of the missing data,

After replacing the missing value, the normalization process is applied to the data set for reducing the dimensionality of the data which used to improve the performance of the proposed system. Then the preprocessed data set is used for further air quality process analysis.

3.2. Predicting pollutants level of concentration

Predicting is the process of identifying the value of new data which is calculated with the help of the predefined set of training data. The proposed system uses the Gradient Boosting based Hierarchical

Temporal Memory Neural Networks (BHTMNN) approach to predicting the pollutant values like nitric oxide (NO), nitrogen dioxide (NO2), sulfur dioxide (SO2), carbon monoxide (CO) and particulate matters in the preprocessed meteorological data. The Temporal Memory Neural Networks is one of the machine learning algorithm [13] which used to predict the pollutant concentration level using the two different modes such as learning and inference mode. The HTMNN network has collection nodes, each node considered as the neurons that forms the tree for predicting the value of the new data. The low level nodes are combined to form the strongest tree structure because, the higher level node only contains the information about the particular value.

3.2.1. Learning mode

Learning is the important step while predicting the air pollutant concentration because, the learning process helps to adapt the meteorological data for identifying the particular new categorical value. It has two important stages, namely spatial pooling and temporal pooling, which uses the cortical learning algorithm. The Cortical learning algorithm is nothing but which stores the every input sequence pattern in memory for the future recognition process. In the spatial pooling stage the number of input possibility is reduced by applying the identifying and memorizing process. During the this process, the input patterns are observed frequently and stored in the neural cortex location for further recognition process. The similarity between the input or data is analyzed and then most similar data's are combined together for making the strongest predicting process which is done with the help of the Gradient Boosting approach [14]. The approach analyzes the dependency between the information by using the following equation 2.

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + \text{const} \quad (2)$$

Where $h_i(x)$ is the similarity between the information which is belongs the same class.
γ_i is search space in the given set of preprocessed meteorological training data set.
The constant value is calculated as follows,

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma) \quad (3)$$

y_i is the output variable of the training data set.
After identifying the data similarity then the residuals have been calculated for grouping the same information into the same class which is done as follows,

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \quad (4)$$

Based on the calculated residuals the grouping and tree forming should be performed. In the temporal spooling, the input patterns are continuously monitored and the related output values are grouped together for matching the patterns in the inference mode and the proposed prediction process depicted in the figure 2.

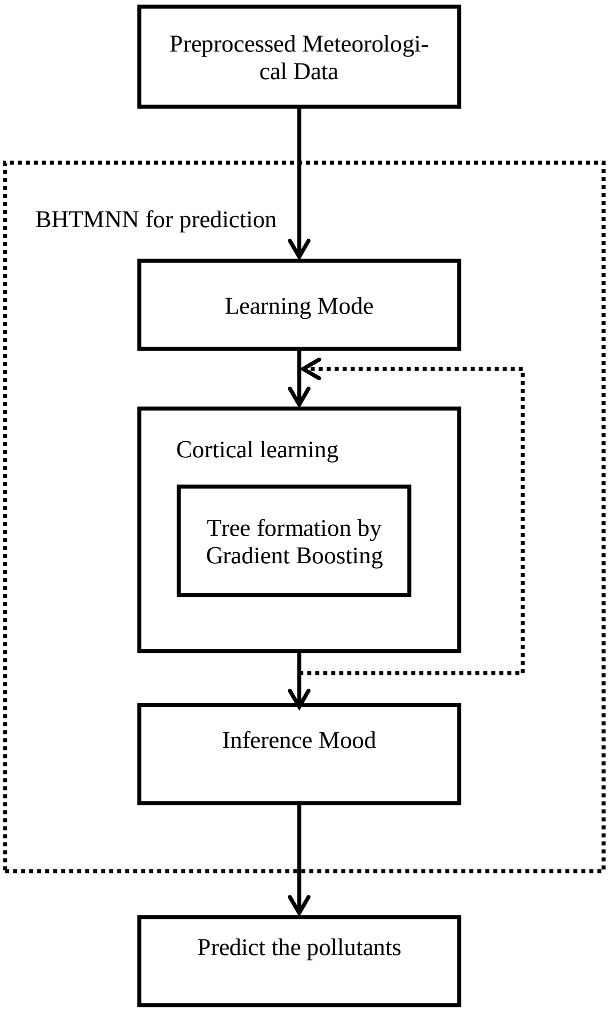

Fig. 2: Pollutants Prediction Process of Meteorological Data.

The above figure 2 explains that the prediction process of the proposed system, the preprocessed data is trained with the help of the learning mode with two different pooling techniques and find the possibilities or similarity between the value and the residual also identified by using the Gradient Boosting method. Then the tree has been formed continuously for all the training data set which contains the particular group of values that used in the time of prediction which is done in the inference mode.

3.2.2. Inference mode

The next stage is the inference mode in which the probability value of the grouped values is calculated and the output is called as the beliefs. From the estimated output value, the temporal tree has been formed by using the following equation 5.

$$h_m(x) = \sum_{j=1}^{J} b_{jm} I\,(x \in R_{jm}) \quad (5)$$

Where, b_{jm} is the predicted pollutant concentration in the region
If the calculated probability values is not belonging to the expected region outcome, then the similar process is repeated for grouping the air pollutants together. Then the incoming input is matched to the group and predict the expected output from the temporal memory tree because the highest level node contains the related output of the particular node. This method simultaneously updates their memory depending on the incoming meteorological data so, it overcomes the existing methods drawback in the training stage itself. The estimated pollutants are used to identify the effects of the air pollution on the human health, which is done with the help of the Fuzzy based CART classification.

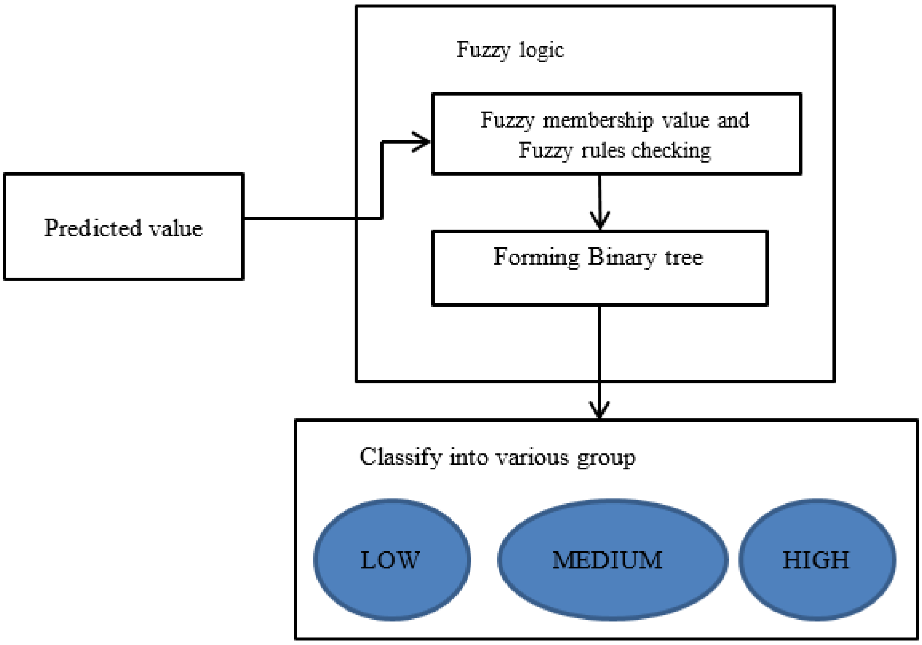
Fuzzy logic
Fuzzy membership value and Fuzzy rules checking
Predicted value
Forming Binary tree
Classify into various group
LOW
MEDIUM
HIGH

Fig. 3: Fuzzy Logic and CART Based Classification.

3.3. Determining the strength of the pollutants using fuzzy based CART

The pollutants strengths are classified by applying the Fuzzy based Classification and Regression Tree approach. FCART approach is one of the supervised classification. approach which forms the set of rules based on the fuzzy. The FCART method provides the strength of the pollutants by forming the decision tree. During the tree forming the fuzzy rules [15] are generated based on the degree of the truths based on the value the pollutants are classified into different groups namely high, low and medium. The fuzzy set forms the rule based on the membership value which means 0 and 1 because the CART [16] approach forms the binary tree while making the decision. Based on the fuzzy value the tree has been formed and pruned the tree depending on the fuzzy value which is used to reduce the cost and complexity while predicting the pollutant strength. The following Figure 3 shows that the classification of predicted values using the fuzzy logic and the classification and regression tree approach.

During the tree formation, when the value is 0 it belongs to low group otherwise the air pollutant fallen into the other group. Then the classification and regression tree reduced the impurities while matching the incoming meteorological data with the help of the information gain and entropy values. The classified group helps to identify the concentration level of the each particle in the air. Each particle in the pollutants affects the human repository system like, the carbon tetrachloride (CCl4) affects the human Stomach, Intestines, Liver and Kidney, SO2 affect the human Eye, Nose and Throat, NO2 affect the lungs and so on. So, the air pollutants which come under the serious category that should avoided by the air pollution control activities. Thus the proposed system identifies the pollutants and strength with low complexity and cost which lead to reduces the air pollution.

4. Results and discussion

Air pollution is the dangerous atmosphere problem because it has several air pollutant substances, like carbon monoxide (CO), sulfur dioxide (SO2), nitric oxide (NO) nitrogen dioxide (NO2) and particulates. So, the proposed system uses the Ariyalur Meteorological data set for predicting the air pollutants and strength by applying the Gradient Boosting based Hierarchical Temporal Memory Neural Network and Fuzzy based Classification and Regression Tree methods. The data has been collected from the Tamilnadu government released information and public website which consists of rainfall level for each month, weather condition, sunny level and so on. (https://www.accuweather.com/en/in/ariyalur/195945/weather-forecast/195945)This method identifies the particles presents in the environment and also estimate the level of concentration which is used to analyze the effects on the human health. Thus the implementation of the proposed system is done with the help of the MATLAB which produces the high prediction rate results. Then the performance of the proposed system is analyzed with the help of the mutual information, mean square error and prediction accuracy,

Mutual Information

The dependency between the actual and the predicted value is estimated by using the mutual information [17] which is measured as follows,

$$MI = \sum_{y\epsilon ybin} \sum_{O\epsilon Obins} p(y, O) \log\left(\frac{p(y,O)}{p(y)p(O)}\right) \tag{6}$$

Where $p(y, O)$ is the joint probability of observation and prediction value.

The following Table 1 shows that the mutual information value of the different prediction techniques such as Hidden Markov Model [18], Backpropagation Neural Networks (BPN) [19], Radial Basis Function Neural Networks (RBFN) [20], Deep Neural Networks (DPN) [21].

Table 1: Mutual Information for Different Prediction Technique

S. No	Prediction Techniques	Mutual Information
1	HMM	0.763
2	BPN	0.812
3	RBFN	0.864
4	DNN	0.746
5	FCART-BHTMNN	0.932

The above table 1 shows that the proposed system has the dependency between the both the observed, predicted value which means, it predicts the air pollutants with high prediction rate

Mean Square Error (MSE)

The MSE [22] value is used to estimate the error present in the pros while predicting the air pollutants which is measured as follows,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \quad (7)$$

y_i is the actual value and y_i' is the predicted value.
The following Table 2 shows that the mean square error of the different prediction techniques such as Hidden Markov Model, Backpropagation Neural Networks (BPN), Radial Basis Function Neural Networks (RBFN),Deep Neural Networks (DPN).

Table 2: MSE for Different Prediction Technique

S. No	Prediction Techniques	Mean Square Error (MSE)
1	HMM	1.56
2	BPN	1.23
3	RBFN	1.05
4	DNN	1.32
5	FCART-BHTMNN	0.221

The above table 2 shows that the proposed system has the minimum mean square error rate, which means, it predicts the air pollutants with high prediction rate. The following figure 4 shows that the Mean Square Error (MSE) of different classification techniques.

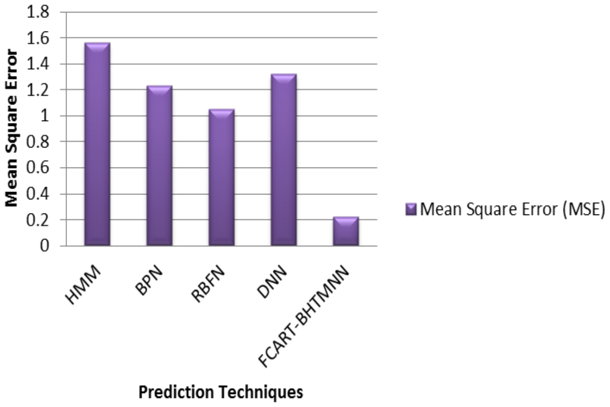

Fig. 4: MSE Error for Different Prediction Techniques.

These reduced Mean Square Error rate helps to increase the performance of the prediction system during the pollutant prediction process. The following Table 3 shows that the predicted rate of different prediction techniques.

Table 3: Predicted Rate of Different Prediction Techniques

S. No	Prediction Techniques	Prediction Accuracy
1	HMM	79.3
2	BPN	84.2
3	RBFN	86.245
4	DNN	81.6
5	FCART-BHTMNN	98.23

Then the following Figure 5 shows that the prediction accuracy of different prediction techniques.

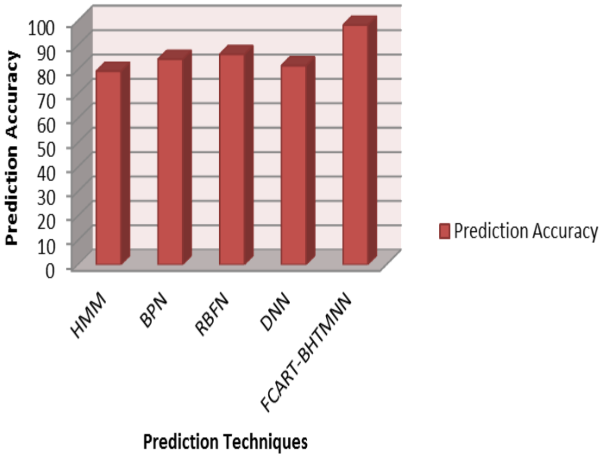

Fig. 5: Prediction Accuracy of Different Prediction Techniques.

Thus the proposed system predicts the pollutants in the air by using the Ariyalur meteorological data with the highest prediction rate. From the classified group, the concentration of the pollutant is identified which will be avoided by several air pollution control activities. Then the performance of the system is evaluated with the help of MSE and prediction rate.

5. Conclusion

Air pollution spoils the atmosphere and affects the human health which lead to increases the death in the world. So, the air pollution is controlled by using the air quality prediction and managing process. Thus the paper analyzes the Ariyalur meteorological data set with the help of the Fuzzy based Classification and Regression Tree and Gradient Boosting based Hierarchical Temporal Memory Neural Networks. These methods train the data set with sequence of memory learning and storage method and predicts the pollutant using the inference mode. Then the concentration of the pollutants is identified by using the fuzzy rules based decision. Finally the pollutants are classified into the groups which used to recognize the severity of the pollutant. Then the performance of the proposed system is evaluated with the help of the experimental discussion.

References

[1] Ioannis N. Athanasiadis and Kostas D. Karatzas and Pericles A. Mitkas, “Classification techniques for air quality forecasting” In Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 17th European, Conference on Artificial Intelligence, Riva del Garda, Italy, August 2006.

[2] Marcelo Arenas, Leopoldo Bertossi, Loreto Bravo, Laura Gallardo, Achim Sydow, “Environmental Information System For Analysis And Forecast Of Air Pollution (Application To Santiago De Chile), available at., http://web.ing.puc.cl/~marenas/publications/icems00.pdf.

[3] R. Shad a, H Ashoori b, N. Afshari b, “Evaluation of Optimum Methods for Predicting Pollution Concentration in Gis Environment”, available at, http://www.isprs.org/proceedings/XXXVII/congress/2_pdf/2_WG-II-2/26.pdf.

[4] Niharika,Venkatadri M,Padma S.Rao, “A survey on Air Quality forecasting Techniques”, International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 103-107.

[5] Mohammad F. Ababneh, Ala’a O. AL-Manaseer and Mohammad Hjouj Btoush, “PM10 Forecasting Using Soft Computing Techniques”, Research Journal of Applied Sciences, Engineering and Technology 7(16): 3253-3265, 2014. https://doi.org/10.19026/rjaset.7.669.

[6] Zhongliang Yue, Yuying Jia, Changqing Zhu, “Prediction of Air Quality During 2010 Asian Games in Guangzhou”, 3rd International Conference onBioinformatics and Biomedical Engineering, 2009. https://doi.org/10.1109/ICBBE.2009.5163212.

[7] Tian, Kadri, Zhang, Feng, Juan, Na, “A Novel Cost-Effective Portable Electronic Nose for Indoor-/In-Car Air Quality Monitoring”, 2012 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring (CDCIEM).

[8] Ana Russo, Frank Raischel , Pedro G. Lind, “Air quality prediction using optimal neural networks with stochastic variables”, Atmospheric Environment in Elesvier, 2013.

[9] A. Suárez Sáncheza, García Nietob, Riesgo Fernándeza,del Coz Díazc, Iglesias-Rodrígueza, “Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain)”, Mathematical and Computer Modelling in Science Direct, Volume 54, Issues 5–6, September 2011.

[10] M. Maruf Hossain, Md. Rafiul Hassan, Michael Kirley, “Forecasting Urban Air Pollution Using HMM-Fuzzy Model”, Advances in Knowledge Discovery and Data Mining, Volume 5012, 2008.

[11] Xiao Feng, , Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, Jingjie Wang, “Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation”, Atmospheric Environment in Elsevier, Volume 107, April 2015.

[12] Kinnari Patel, Mehta, Raghuvanshi, “Incremental Missing Value Replacement Techniques for Stream Data”, International Journal of Computer Applications (0975 – 8887) Volume 122 – No.17, July 2015.

[13] Nguyen, Starzyk, Wooi-Boon Goh, Jachyra, "Neural Network Structure for Spatio-Temporal Long-Term Memory", IEEE Transactions on Neural Networks and Learning Systems, 2012. https://doi.org/10.1109/TNNLS.2012.2191419.

[14] Khot, Natarajan, S. Kersting, Shavlik, "Learning Markov Logic Networks via Functional Gradient Boosting", International Conference on Data Mining (ICDM) in IEEE, 2011

[15] Om Prakash Verma, Himanshu Gupta, "Fuzzy Logic Based Water Bath Temperature Control System", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012.

[16] Wei-Yin Loh, "Fifty Years of Classification and Regression Trees", International Statistical Review (2014), 82, 3, 329–348 https://doi.org/10.1111/insr.12016.

[17] Jim Jing-Yan Wang , Yi Wang , Shiguang Zhao , Xin Gao, "Maximum mutual information regularized classification", Engineering Applications of Artificial Intelligence in Elesvier, 2015.

[18] Artemio Sotomayor-Olmedo, Marco A. Aceves-Fernández, Efrén Gorrostieta-Hurtado, Carlos Pedraza-Ortega, Juan M. Ramos-Arreguín, J. Emilio Vargas-Soto, "Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A Kernel Performance Approach", International Journal of Intelligence Science, 2013, 3, 126-135 https://doi.org/10.4236/ijis.2013.33014.

[19] Mouhammd Alkasassbeh, "Predicting of Surface Ozone Using Artificial Neural Networks and Support Vector Machines", International Journal of Advanced Science and Technology, Vol. 55, June, 2013.

[20] Maurizio Caselli, "A simple feed forward neural network for the PM10 forecasting: comparison with a radial basis function network" available at., http://new.sis-statistica.org/wp-content/uploads/2013/10/CO09-A-simple-feed-forward-neural-network-for-the-PM10.pdf.

[21] Bun Theang Ong, Komei Sugiura, Koji Zettsu, "Dynamically pretrained deep recurrent neural networks using environmental monitoring data for predicting PM2.5", Neural Comput & Application in Springer, 2015.

[22] S. Galmarini, I. Kioutsiouki and E. Solazz, "E pluribus unum: ensemble air quality predictions", Atmospheric Chemistryand Physics in 2013.