Air Pollution Prediction in Smart Cities using Artificial Neural Network

Venkata Taghu Ram R[a], Saiteja Royal Cheruvu[b], Venkata Shashank Gokul G[c], Surya Tataji Tanala[d]

[a] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

[b] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

[c] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

[d] School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India

Abstract:- The rapid development of urbanization and industrialization led to the suffering of many developing countries from heavy air pollution. The rowing concern for air pollution has been raised by many governments and people worldwide because it affects human health and sustainable development. Particularly, in India, the drastically deteriorating air quality threatens the health of its people. Meanwhile, in smart cities, knowledge of timely and reliable levels of air pollution is essential for the effective set-up of smart pollution systems. The method uses the Artificial Neural Network-Linear Vector Quantization (ANN-LVQ). It is an integration of the NN with LVQ, a significant Air Prediction Technique. The main aim of implementing this method is to provide early warnings by predicting air quality and estimate the influencing pollutant that contaminates the quality of air which thereby leads to air pollution. The performances of the methodologies proposed for this model is assessed in terms of accuracy, precision and recall.

Keywords:- ANN, LVQ, NN, Accuracy, Recall, Precision

I. INTRODUCTION

Air (containing oxygen) is an indispensable necessity for supporting life. At first, on the air of the earth, there was no oxygen and, surprisingly, then, at that point, the life was there. Gradually the oxygen requiring organic entities created with the advancement of oxygen and at present there is around 21% oxygen in the air and pretty much every animal consumes it for its digestion. Luckily, nature has the course of photosynthesis wherein the green plants produce oxygen (O_2) within the sight of daylight and consume carbon-dioxide (CO_2) [1, 2]. In this way the convergence of O_2 is recovered and the CO_2 delivered by the human and normal exercises is consumed. CO_2 as such is definitely not a destructive gas. It behaves like glass in the nursery impact and hence keeps the earth warm. It permits the sun shortwave radiation to go through it and doesn't permit the earth longwave heat radiation to get away from through it.

Consequently it keeps the earth warm. On the off chance that there had been no CO_2 on earth it would have been not livable. The temperature of the world's surface would have been negative without any it. With the goal that way it is a helpful gas. A dangerous atmospheric deviation implies an expansion in the temperature of the earth bringing about over softening of ice and rising water level in the seas and an expansion in vector borne sicknesses and the distress [3].

This is the idea of Pollution. Contamination implies blending of anything unfortunate, prompting some unsafe impact. In the event that CO_2 is there, a specific level of CO_2 is attractive to keep the earth warm and comfortable, however past that, it expands the temperature so high which is bothersome and consequently causes air contamination. The major cause of air pollution accounted for is vehicular pollution due to increasing population and lack of preventive measures in this direction. From 2007 to 2017, around 1.40 lakh vehicles were registered in Jaipur and 9.8 crore vehicles were registered in Rajasthan 74.24% of them two-wheelers and almost all of them two-stroke [4, 5].

II. RELATED WORK

Several Machine Learning and Artificial Neural Network techniques have been introduced to resolve issues on air pollution prediction and forecasting in the environment of smart cities. However, to the best of our knowledge, very few systematic literature reviews have been performed on the use of ANN in predicting air quality for smart cities. Most urban areas all over the planet have routine meteorological stations that action fundamental meteorological boundary. Factors, for example, limit layer profundity and soundness records are not promptly accessible in that frame of mind, regardless of being

significant for the scattering cycles of barometrical contaminations [6, 7]. Nitty gritty data about the strength of emanation sources is restricted to a couple of urban communities. Thusly, more examinations are expected to foster air quality models that can enough catch the changeability in quantifiable focuses utilizing restricted information on the meteorological qualities and discharge boundaries. The capacity of ANN procedures to catch the nonlinear way of behaving of complicated air processes makes it a reasonable instrument for growing such models. In particular, multi-facet perceptron ANN models have been generally utilized in air sciences for expectation, capability estimation, and example characterization lately [8].

In this study, a comprehensive literature review was performed to identify research activities that led to the development of a hybrid pollution control system. To effectively perform this literature survey, research papers published in leading databases that were related to this subject were examined and briefly described in this chapter. For the systematic literature review process, the research papers were collected and organized before the analysis of parameters and synthesis of the research data. The most commonly used algorithms identified in the systematic literature review comprise the following the fig. 1.

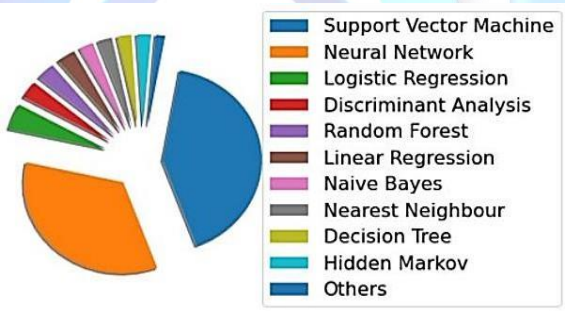

Fig. 1: Algorithms Used

A few strategies for the expectation of air quality, going from mathematical, numerical, and factual techniques (e.g., relapse) to computerized reasoning based procedures like ANN, have been distinguished in the writing studies. Subsequently, the most effective way to accomplish the sufficient air contamination examination results is to utilize the brain network models related to an equipped methodology of the ecological Management [9]. Every one of the meteorological factors and factors have a non-direct relationship with air quality, which can be precisely caught by nonlinear models like ANNs. A brief overview of each study is outlined as shown below

Wavelet Transformation and ANN were employed to predict the daily routine concentration of particulate matter using neural network predictors along with the linear ARX model. To ensure the predictive accuracy of the entire quality inspection technique, various trials were used to predict PM10 pollution in Warsaw to produce acceptable results [10].

ANN techniques were applied using periodic Air Pollution Index (API) datasets from industrialized and domestic monitoring stations in Malaysia to forecast the API values for the past ten years. The results of each technique were compared with the Root Mean Square Error (RMSE).The study concluded that ANN can offer a more reliable approach as compared to FTS and ARIMA [11].

It employed a two-step approach to predict pollutant concentrations with high spatio-temporal resolution. NN models were initially used to generate short-term temporary forecasts for predicting air pollution and weather forecast data, and thereafter, these forecasts were later compared with previous autonomous data. It was concluded that the prediction of pollutants can be easily performed by implementing the spatial-temporal method. The review results demonstrated that the proposed strategy offers a decent substitute for the characterization of air quality in metropolitan urban communities [12].

III. ANN

The statistical models that have a functional similarity to a human brain are Artificial Neural Networks Modern technologies including Voice Recognition, Image Recognition, and Robotics are the recent advancement in the field of ANN. A neuron is the basic functional unit of the human brain and the term 'Brain' is gotten from the human (creature) sensory system. Like the Human mind a brain network gains information through learning and this information put away inside the synaptic weights to generalize and derive a functional relationship for further processing [13].

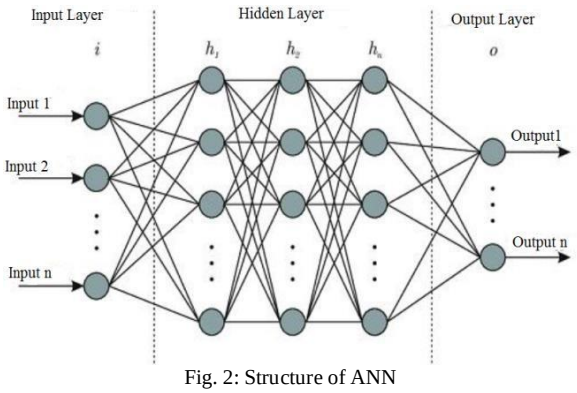

Fig. 2: Structure of ANN

Multiple layer perceptron as the name represents has multiple hidden layers along with input and output layer. A simple neural network has three layers is represent in fig. 2.

Input layer – To process data into the system of neural network this is the beginning. The information gathered from the outer world fed up into the system through this layer. The selection of input independent features generally depends on the type of problems and correlation among variables. Several new studies have been carried out to select the most significant features that may affect the desired output.

Output layer – this layer has nodes which represents the results or information which we want to get from the input data and the output signals have somehow correlated with input signals.

Hidden layer – The information fed into the system forward to the subsequent hidden layers which lies between input and output layers. These layers are interconnected to the system with defined weight and activation functions. After learning these layers produce functional relationship between input and output signal to find desired output as dependent variables.

Advantage-

The reason artificial neural network preferred over other traditional and semi-empirical models is less computational time and its understanding for complex relationships [14, 15].

- Their complex problem solving capabilities with high prediction accuracy is well known.
- Dealing with missing and noisy data as well as can handle large number of variables is outstanding. As selection of parameters can increase the model prediction efficiency and accuracy.
- It has incredible learning and generalization capabilities and does not need to be reprogrammed.
- It can be implemented in numerous practical as well as theoretical problems and has wide variety of applications.

IV. METHODOLOGY

ANN is a mathematical model that facilitates the organization and/or functional aspect of a biological NN.ANN, as a rule, is a specialized system that allows an organization to rely on external or internal information that crosses the network in the learning process. The model often complexes the input and output relationships or use them to neutralize patterns in the data. Neurons are often grouped into three types of layers. The input level is a set of neurons that receive input through a user agent. The output level is the level of the neuron that sends data to the user agent. Whereas, between the two levels are the hidden layers. Fig. 3 illustrates the architecture of the ANN.

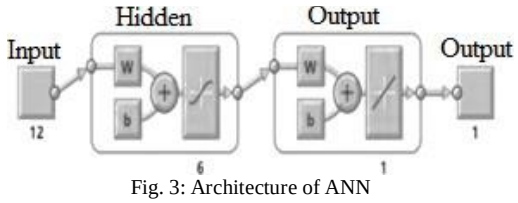

Fig. 3: Architecture of ANN

The hidden layer neurons are common to other neurons and do not interact directly with a user agent. Each neuron can influence processing that can carry a certain level of NN. However, hidden layers are optional in NNs. While the input and output levels are important, where the layers can serve as both entry and exit level. The NN determines the classified groups by using many epochs as training progresses. Training is how these connection weights are defined. Most training cycles start by marking the random numbers to the weight gain. In this proposed method, weighting parameters are selected by using LVQ technique for prediction. Back propagation calculation is used as part of the ANN feed forward. Input and output levels consist of nodes corresponding to input and output variables, respectively. Meanwhile, data move between these layers through weighted connections. Fig. 4 presents the flow chart of the proposed ANN-LVQ.

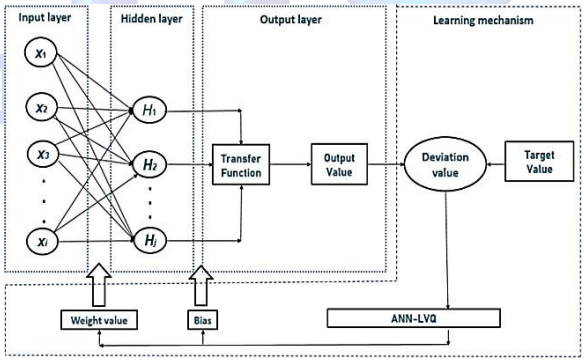

Fig. 4: Model of ANN-LVQ

We node receives data from the previous level and calculates the weight of all its input data as mentioned in Eq.(1)

$$x_i = \sum_{j-1}^{n} \Delta W_{i,j} X_j \quad \textbf{(1)}$$

Where the number of inputs is defined as n, the weight of the connection between nodes i and j is represented as ΔW, meanwhile, input from node j is

defined as X. In this equation, the LVQ technique is used to choose ΔW.

V. SIMULATION RESULT

The accuracy rate alludes to a suitable evaluation index to assess the efficiency in AQI prediction and recommendation, while the error rate alludes to the lack of accuracy of the prediction in the output variable. In short, the accuracy rate refers to the percentage of the total number of correct predictions, whereas the error rate refers to the percentage of the total number of incorrect predictions.

Table 1 represents the comparisons between the performances of the ANNLVQ model, existing ANN and machine learning models. Fig. 5, fig. 6 and fig. 7 is the graphical representation of the proposed ANN-LVQ model over the existing and machine learning model.

Table 1: Comparison Result

Methodology	Accuracy	Precision	Recall
Machine Learning	67.82 %	58.32 %	64 %
Existing ANN	84.56 %	73.54 %	83.11 %
ANN-LVQ	92.34 %	86.74 %	89.42 %

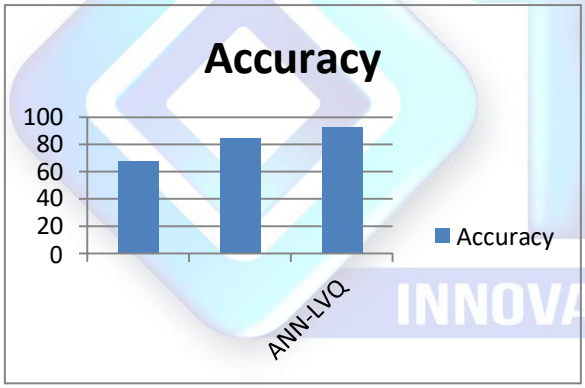

Fig. 5: Graphical Accuracy

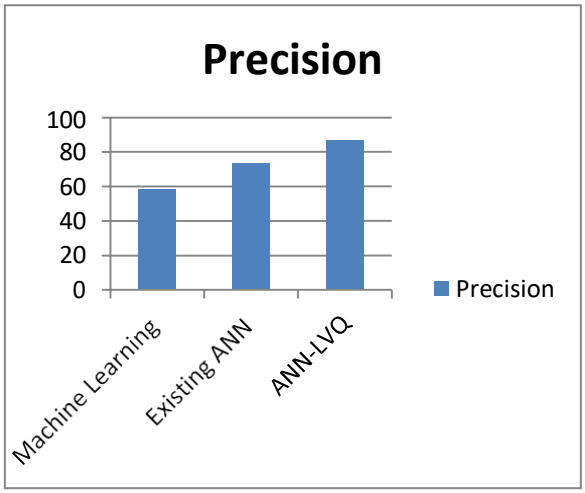

Fig. 6: Graphical Precision

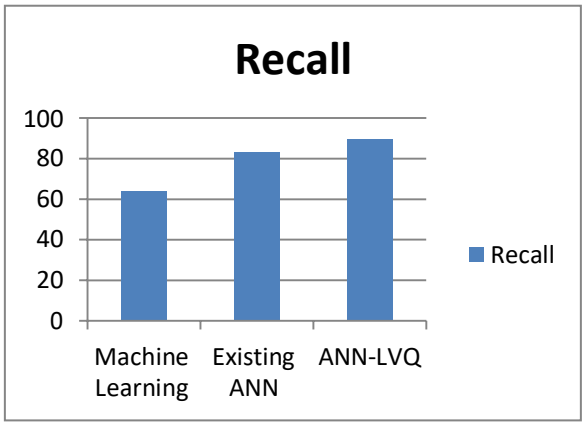

Fig. 7: Graphical Recall

For this section, only major cities namely Indore, Pune and Delhi were selected because of the high variation of air pollution level in these selected cities among other cities. Table 2 indicates the predicted parameter values of ANN-LVQ for the three major cities. And graphical represent for major cities is shown in fig. 8-10.

Table 2: Results for Major Cities

Cities	Methods	Accuracy	Precision	Recall
Indore	Machine Learning	72%	56.4%	63%
	Existing ANN	79%	63.2%	68.2%
	ANN-LVQ	84.7%	74%	78.6%
Pune	Machine Learning	69.2%	57%	58%
	Existing ANN	77.3%	59.8%	69.3%
	ANN-LVQ	83.5%	67.6%	72.4%
Delhi	Machine Learning	71%	52%	57.4%
	Existing ANN	76%	62.3%	61.4%
	ANN-LVQ	88.4%	77.3%	71.4%

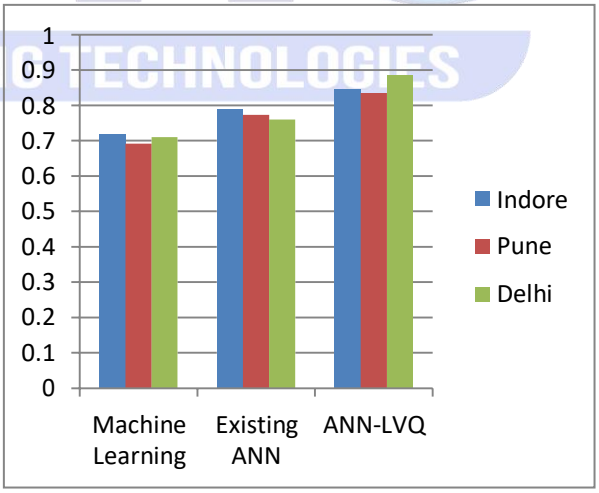

Fig. 8: Graphical Accuracy

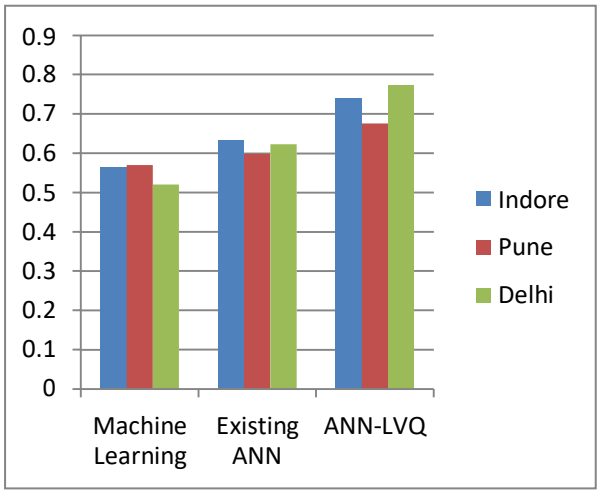

Fig. 9: Graphical Precision

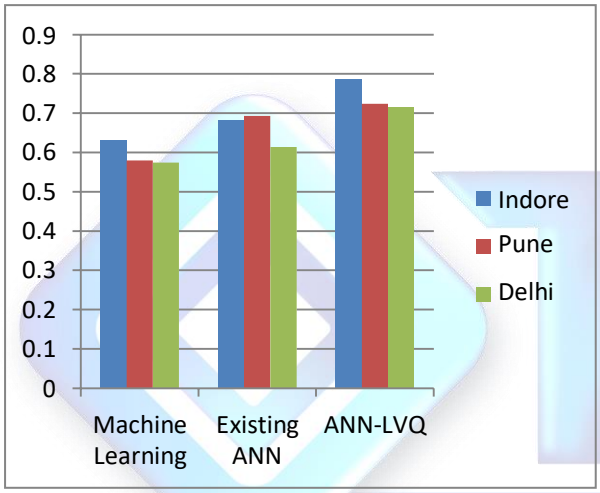

Fig. 10: Graphical Recall

VI. CONCLUSION

It can be concluded that predicting air quality is a multifaceted task because of its dynamic characteristics, volatility and spatial-temporal variability of pollutants and particles in the air. Moreover, due to the critical impact of air pollution on the population and environment, the prediction and tracking of air quality, especially in urban areas are gradually changing. The results yielded by the proposed model indicated that the fusion model of ANN is still effective in detecting air quality and predicting the pollution level in urban, rural or industrial regions. The weighting function of the implemented ANN is optimized for the collected data set through LVQ. In total, six categories were predicted from the cleaned data and were used to predict AQI. ANN verification with LVQ was examined on datasets from different cities in India from 2015 to 2020. The accuracy of the proposed model is significantly higher than the other existing research models. Hence, this model can be effectively applied to the data collected from other smart cities for air pollution study.

REFERENCES

[1] S. M. L. S. Cabaneros, J. K. S. Calautit, and B. R. Hughes, "Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO2 Pollution," in Energy Procedia, 2017, vol. 142, pp. 3524–3530.

[2] S. Sanober and K. Usha Rani, "Review on neural network algorithms for air pollution analysis," in Advances in Intelligent Systems and Computing, 2020, vol. 1054, pp. 353–365.

[3] L. Bai, J. Wang, X. Ma, and H. Lu, "Air pollution forecasts: An overview," International Journal of Environmental Research and Public Health, vol. 15, no. 4. MDPI AG, Apr. 17, 2018.

[4] K. Mahesh Babu and J. Rene Beulah, "Air quality prediction based on supervised machine learning methods," Int. J. Innov. Technol. Explor. Eng., 2019.

[5] Mo Z, Huang J, Chen Z, Zhou B, Zhu K, Liu H, Mu Y, Zhang D, Wang S. Cause analysis of pm 2.5 pollution during the covid-19 lockdown in Nanning, China. Sci Rep. 2021;11(1):1–13.

[6] Zoran MA, Savastru RS, Savastru DM, Tautan MN. Assessing the relationship between surface levels of pm2.5 and pm10 particulate matter impact on covid-19 in Milan, Italy. Sci Tot Environ. 2020;738:139825.

[7] Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and covid-19 infection: evidence from China. Sci Tot Environ. 2020;727:138704.

[8] Lin K-P, Pai P-F, Yang S-L. Forecasting concentrations of air pollutants by logarithm support vector regression with immune algorithms. Appl Math Comput. 2011;217(12):5318–27.

[9] Zaidan MA, Dada L, Alghamdi MA, Al-Jeelani H, Lihavainen H, Hyvärinen A, Hussein T. Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies. Appl Sci. 2019;9(20):4475.

[10] Garg S, Jindal H. Evaluation of time series forecasting models for estimation of pm2.5 levels in air. In: 2021 6th international conference for convergence in technology (I2CT). IEEE; 2021, p. 1–8.

[11] M. Pawul and M. Śliwka, "Application of artificial neural networks for prediction of air pollution levels in environmental monitoring," J. Ecol. Eng., vol. 17, no. 4, pp. 190–196, 2016.

[12] S. Ameer et al., "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities," IEEE Access, vol. 7, pp. 128325–128338, 2019.

[13] A. Challoner, F. Pilla, L. Gill, G. Adamkiewicz, and M. P. Fabian, "Prediction of Indoor Air Exposure from Outdoor Air Quality Using an Artificial Neural Network Model for Inner City Commercial Buildings," mdpi.com, 2015.

[14] W. Ding, J. Zhang, and Y. Leung, "Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks," Environ. Sci. Pollut. Res., vol. 23, no. 19, pp. 19481–19494, Oct. 2016.

[15] N. H. A. Rahman, M. H. Lee, Suhartono, and M. T. Latif, "Artificial neural networks and fuzzy time series forecasting: an application to air quality," Qual. Quant., vol. 49, no. 6, pp. 2633–2647, Nov. 2015.