

## Simulation and optimization of genetic algorithm-artificial neural network based air quality estimator

Shirish Pandey, S.Hasan Saeed, N.R. Kidwai

Department of ECE, Integral University, India

---

### Article Info

#### Article history:

Received Jan4, 2020

Revised Mar6, 2020

Accepted Mar21, 2020

---

#### Keywords:

Artificial neuralnetwork

Genetic algorithm

---

### ABSTRACT

In this work intelligent model for estimation of the concentration of carbon monoxide in a polluted environment is developed on mat Lab platform. The results are validated using data collected from repository linked to University of California. The data records are over the duration of one year using E nose sensor placed in main city of Italy. The records are rectified and segmented at different length to extract the Base and Divergence Values features. An Artificial Neural Network Model (ANN) is developed and the result is validated manually. Another optimized Genetic Algorithm-Artificial Neural Network based air quality estimation model is developed which validate the result using artificial intelligence technique to get a better performance network.

Copyright © 2020 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

ShirishPandey,  
Department of ECE, Integral University,  
Kursi Road Lucknow 226021, India.  
Email: Shirish109@gmail.com

---

## 1. INTRODUCTION

The environment air pollution affects the human life. Therefore it is necessary to predict the air quality exactly. In the present time it will be rather easy to use digital signal processing technique for research purpose and so far generating result. The geographical information system (G.I.S.) a power technique to process polluted data. Though the G.I.S based data mining technique is very popular but it is not much accurate. People involved in environmental research also use a number of numeric factors to achieve desired result. It is easier to define certain attributes for each data sample using numeric factors. Due to these attributes it is possible to identify variation in pollution. The attributes are ordinal-scale variable. Ordinal-scale variable are obtained by literature review. Ordinal scale variable are not easy to define [1]. A number of other methods were also proposed for the estimation and monitoring air quality. A method was proposed to monitor air quality in environment using concept that in a vehicle the air quality is similar to the outer environment if the windows are open. They detected air exchange rate within and outside of vehicle before and after window is open on an internet of things (Iot) platform [2]. Another IoT application using ANN is proposed by overcoming the limitation of existing approaches which have lengthly short term memory based artificial intelligence technique using neural network for prediction of contaminants concentration in a polluted environment [3-5].

Different methods for monitoring environment quality in manufacturing industry using wireless sensor network with Zigbee were proposed to collect data for various toxic gases along with temperature, and humidity [6-11]. A unique method to predict the air quality using neural network based on multiple locations, outcome from correlations between nearby locations and among similar locations in temporary domain at the location in Taiwan and Beijing [12]. An electronic nose method used to predict the environmental pollution. They employed three type of e-nose system based on amperometric gas sensors [13-15]. A method using wireless sensor network is presented using deployment of sensors at various

location that senses weather condition on time varying frame and infrastructure connectivity [16]. Another different type of approach presented a method for the estimation of air pollutant by neural network to distinguish natural images into separate categories based on PM<sub>2.5</sub> concentration [17]. A system comprises of low cost motes made up of an array of sensors to monitor air quality is developed. Motes are wirelessly linked with sensing platform to process the informative data and forecast the air quality [18-20]. In order to achieve more accurate result the concept of newer and powerful data mining technique such as Artificial Neural Network (ANN) and Genetic algorithm based optimized search are used for Pollution based research [21-23]. A multiple inbuilt sensor network circuitry is proposed to discriminate between PM<sub>2.5</sub> and PM<sub>10</sub> level and also to monitor different air polluting agents [24]. A method which includes double channel weighted convolution network based learning algorithm is proposed for the measurement of air quality [25]. A method which includes machine learning artificial neural network based is proposed for accurate prediction of air quality with metrological measuring system [26]. A unique method which shows that it becomes possible to exactly predict PM 2.5 concentration within minimum computational time by using different data cluster technique on IOT (internet of thing) platform [27].

For the estimation of the urban air quality a method is proposed using vehicular sensor network and a smart data collecting and estimation mechanism is developed. On VSN (vehicular sensor network) platform [28]. The result obtained by ANN model is nothing but logical response based on history of the data. The artificial intelligence technique using Genetic algorithm (GA) on Artificial Neural network platform is also preferred by many researchers [29-34]. The genetic algorithm has important property of genetic inheritance which made it a complex algorithm for data mining. In this research work Genetic Algorithm-Artificial Neural Network based optimized search is use to estimate the concentration of Carbon monoxide in a polluted environment the results so obtained is compared with the Artificial Neural Network Model (ANN) model to justify that Genetic Algorithm-Artificial Neural Network (GA-ANN model) is more accurate.

## 2. PROPOSED RESEARCH METHOD

### 2.1. Artificial neural network

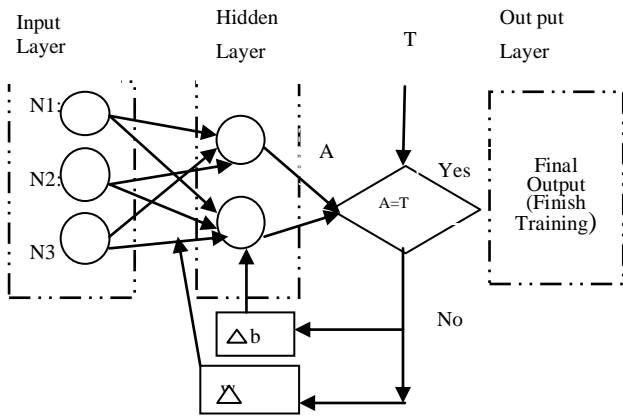
Artificial neural network (ANN) is based on human neural response to receive external stimuli and produce suitable response. For research work it is like to receive external stimulation and then through a number of logical processes produce the outcome. It is to decide which response is most suitable. In general ANN is a method of classification can be demonstrated in many ways [35,36]. By the process of classification the factor effect can be analyzed and to build regression model for predictive simulation. The performance of ANN depends upon sample classification outcome. If the sample description given by some factors can be classified by ANN itself in certain category further the classification meets the reality performance then it is considered as well performance of ANN. The algorithm involved in ANN varies the weight value on every factor to meet the desired classification. This is known as ANN training or learning phase.

ANN is made up of three layers input layer, hidden layer and output layer as shown in Figure 1. Input layer has some input variables which are based on assumption of factors of research work. Output layer is classification of result finally. The algorithm involved in ANN works on classification of samples through hidden and input layer. Hidden layer has much importance than other two layer. ANN is a kind of software model that deals with the given problem in linear or nonlinear system; the hidden layer is not required in linear system. The well performance ANN can also be found by the setting of variable in input layer and then adding hidden layer nodes to achieve result. It is difficult to examine the relationship between ANN classification result and variables, this is why this network model known as black box [37]. In linear network system it is easier to clarify the relation between ANN output and input variable. In this paper we apply the ANN model as non liner system. The weight values of ANN can clarify which factors are more effective in liner ANN system.

### 2.2. Genetic algorithm

Genetic algorithm (GA) is an artificial intelligence technique the main objective is optimized searching. GA is popular because of the global optimization technique and permit synchronous operation of a number of numerical targets. The prime difference comparing with other optimization algorithm is that GA can pass the information from the last stage of search to the next stage, shows the concept of inheriting. A number of research workers apply GA based optimization search into ANN process to find the optimized parameter goal of ANN. This differentiate genetic algorithm from other technique For its global searching ability and to minimize the concept of ANN result sinking into local optimization. A number of research experts have explicitly demonstrated how to use to practice GA in optimized search applications

[38,39].Genetic algorithm involves many processing steps as shown in Figure 2.The main steps are criteria for initial population, population selection, population crossover and population mutation, and then enter into next generation population. Selection step can opt for best objects to conduct genetic processes according to the optimization goal. The selection process can retain the elite objects to next generation. Population Crossover is just mixing of inheritance property for next generation output. In Mutation process some of the inheritance property object of few objects is varied by random searching. Crossover and mutation steps differentiate the GA algorithm from other technique so that genetic algorithm wouldn't limit to the local optimization. Figure 2 shows the basic flow chart of the genetic algorithm.



N: input node, b: Bias, w: weight, T: Target, a: Actual output

Figure 1. Artificial neural network model

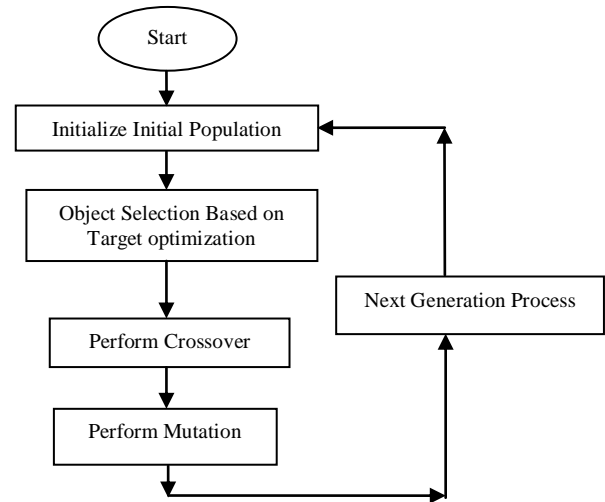


Figure 2. Genetic algorithm flow chart

**3. RESEARCH METHOD**

The objective of this research work to estimate the concentration of carbon monoxide in a polluted environment. The Polluted data is taken from university of California web site which is one year data of the main city of Italy collected at an interval of one hour periodically for one year. The Air Polluted data is Rectified and Modified into Different Segment Length. The Base and Divergence Values are sorted. The Artificial Neural Network Model is Developed Using Mat Lab. Number of Neurons in Hidden Layer along with Data Segment Length is varied to Achieve Best Performance Network. Performance Analysis is done in Terms of Minimum Value of Mean Squared Error and Maximum Value of Regression. After the Identification of Best Performance Network model which is for Segment Length=40 and no. of Neurons in Hidden layer=4, the Concentration of Carbon Mono Oxide is Analyzed on this Network Model. To achieve more accurate result, Genetic algorithm based optimize search coding is done on Mat Lab and is applied on Artificial Neural Network (ANN) network model. Mat LAB Coding of Genetic AlgorithmArtificial Neural Network, (GA-ANN) Model is developed.

**4. RESULT AND DISCUSSION**

Result Generation and Result Analysis is Done for GA-ANN Model. Comparison of ANN Model and GA-ANN Model is done. It is found that GA-ANN model is more accurate First of all the ANN is applied individually at different segment length and different number of hidden nodes. Since ANN exhibits behavior of random updates in network parameters thus it is not sure that in single attempt best network can be produced. Thus keep in this in account 5 attempts for each combination of data segment length and network hidden nodes are performed. The best MSE values obtained at different attempts for definite combination of L and n. It is observed that when ANN is applied at supervised attempts the Minimum MSE obtained is 0.063 at L=40 and n=4.Thus large ANN is not necessarily required. Small number of nodes is more capable of giving better accuracy. After this ANN is hybrid with GA for developing an automated ANN model design approach. This GA-ANN approach search for best {L, n} value at which minimum MSE is obtained (see Figure 3). In Table 1 minimum MSE is obtained using GA-ANN at {L, n} = {161, 18} having value 0.044 which is better than when ANN is applied alone without GA.Table 2 shows the ANN based regression result which express the similarity between expected output and actual output which comes out to maximum 0.86. Table 3 shows the Genetic Algorithm-Artificial Neural Network Based regression result which is 0.95 at its maximum for {L, n}={161,18}. Table 4 shows the

comparison of Regression between ANN and GA-ANN model. It is clear that Regression Based on GA-ANN model is better. Figure 3 shows the detailed arrangement of data processing and sampling/optimization of GA-ANN in terms of flow chart. Figure 4 shows the regression curve. Figure 5 and 6 shows the plot of MSE at different attempts for ANN and also for GA-ANN. Figure 7 and 8 are similar plots but for Regression values. Figure 9 is the comparative plot (mse) for all the attempts for ANN and GA-ANN approach. Figure 10 is the comparative plot (Regression) for all the attempts for ANN and GA-ANN network modeling. Figure 11 shows the variation in fitness value i.e. mean squared error and individual L and n with respect to individual i.d. During network modeling it can be easily observed that GA-ANN always gives lower value of Mean squared error and higher value of regression as compared to attempts when ANN applied alone

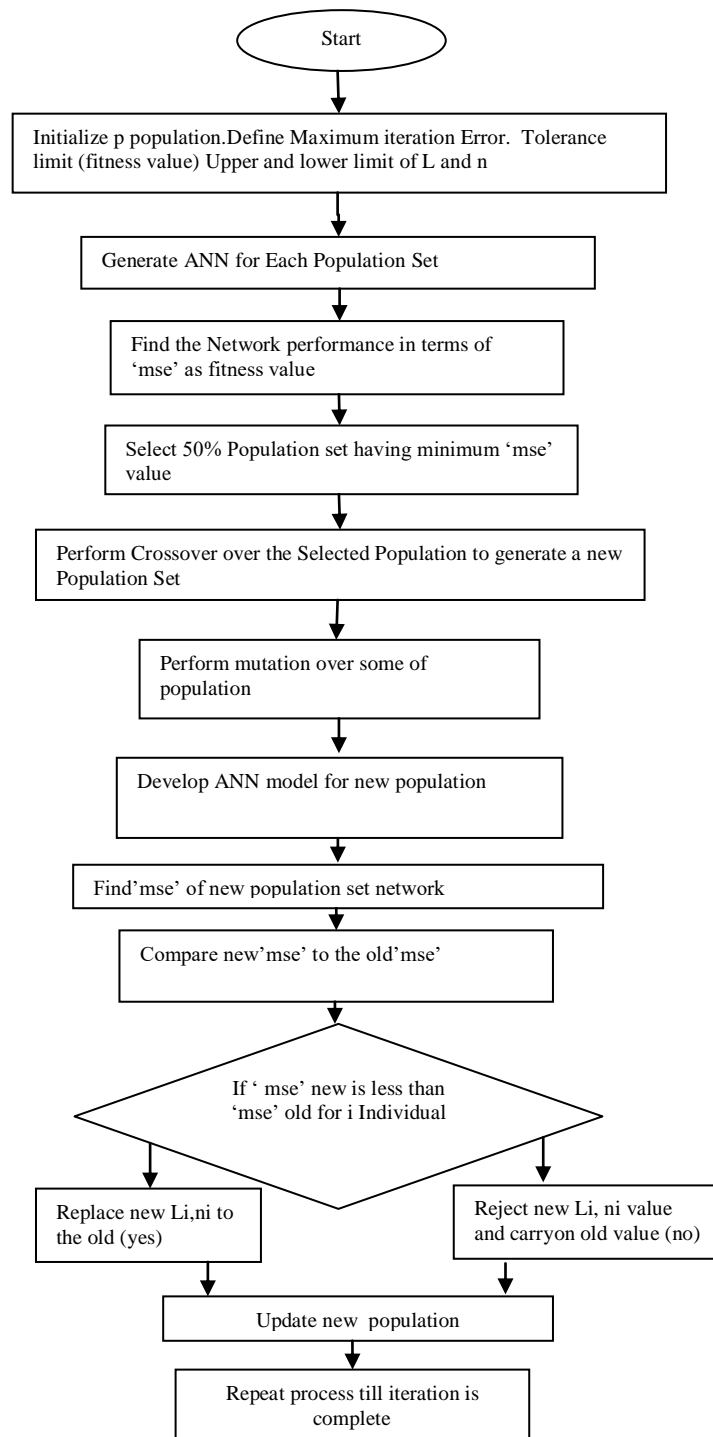


Figure 3. Genetic algorithm-artificial neural network model flow chart

Table 1. Genetic algorithm –artificial neural network based mean squared error comparison

Genetic Algorithm-ANN based Mean square error Comparison			
	Mean Squared Error(Mse)	Segment length(L)	No. of Nodes(N)
ANN	0.131	10	8
	0.1113	15	10
	0.1182	20	2
	0.1178	25	16
	0.1017	30	8
	0.1063	35	14
GA ANN	0.063	40	4
	0.0999	45	8
	0.0444	161	18
	0.0625	153	15
	0.0588	107	14
	0.0612	149	18
	0.0652	158	11
	0.0588	215	13
	0.0517	126	12
	0.0465	170	12
0.0444	161	18	
0.0625	153	15	
0.0588	107	14	

Table 2. ANN based regression data

ANN Based Result			
Attempts	Regression(R)	Segment Length(L)	No of Neurons(n)
B1	0.68	10	8
B2	0.74	15	10
B3	0.72	20	6
B4	0.73	25	16
B5	0.79	30	8
B6	0.77	35	14
B7	0.86	40	4
B8	0.77	45	8

Table 3. GA-ANN based regression data

Genetic Algorithm ANN Based Result			
Attempts	Regression(R)	Segment Length(L)	No of Neurons(n)
A1	0.85	144	14
A2	0.86	214	13
A3	0.89	118	9
A4	0.88	89	9
A5	0.95	161	18
A6	0.89	133	16
A7	0.9	144	14
A8	0.92	214	13

Table 4. Genetic algorithm-artificial neural network based regression data comparison

GA-ANN Comparison				
	Attempts	Regression(R)	Segment Length(L)	No of Neurons(n)
ANN	B1	0.68	10	8
	B2	0.74	15	10
	B3	0.72	20	6
	B4	0.73	25	16
	B5	0.79	30	8
	B6	0.77	35	14
	B7	0.86	40	4
	B8	0.77	45	8
GA-ANN	A1	0.85	144	14
	A2	0.86	214	13
	A3	0.89	118	9
	A4	0.88	89	9
	A5	0.95	161	18
	A6	0.89	133	16
	A7	0.9	144	14
	A8	0.915	214	13

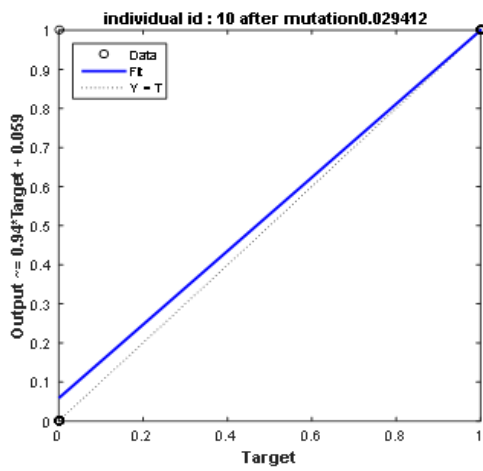


Figure 4. Regression graph

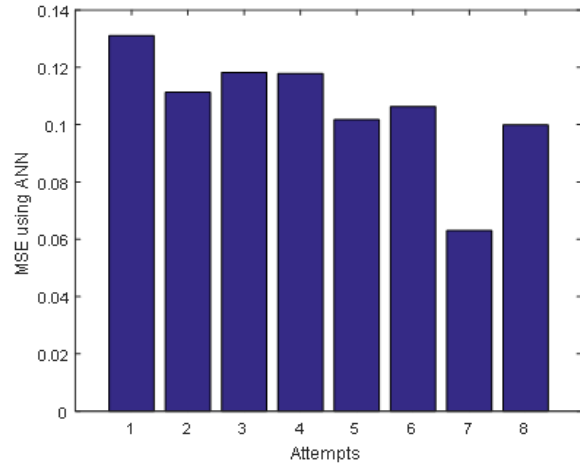


Figure 5. Mse (ANN) with respect to no. of attempts graph

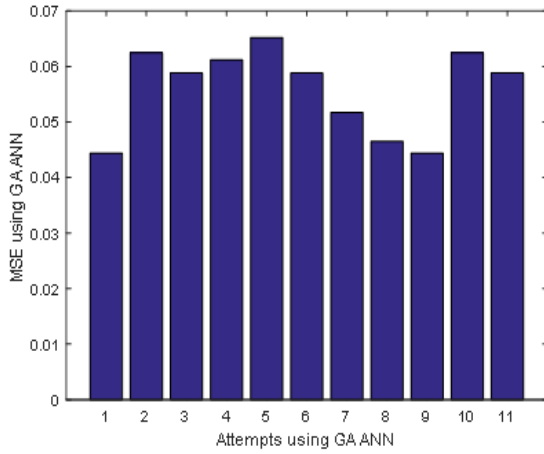


Figure 6. Mse (GA-ANN) with respect to no. of attempts graph

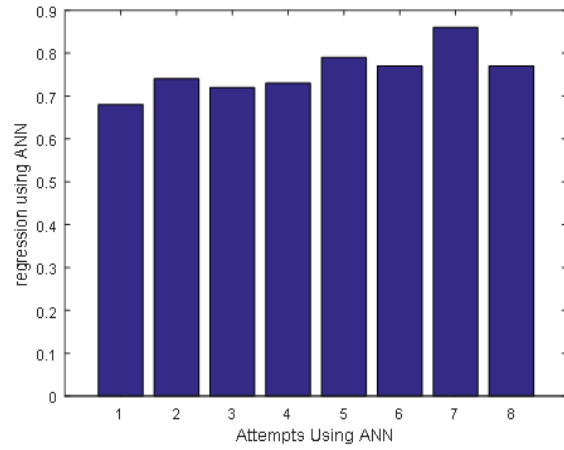


Figure 7. Regression (ANN) with Respect to no. of attempts graph

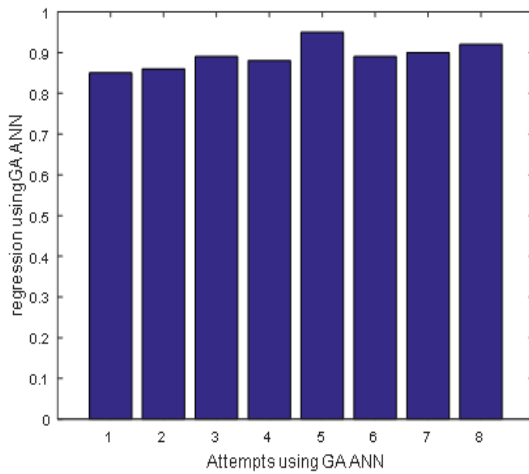


Figure 8. Regression (GA-ANN) with Respect to no. of attempts

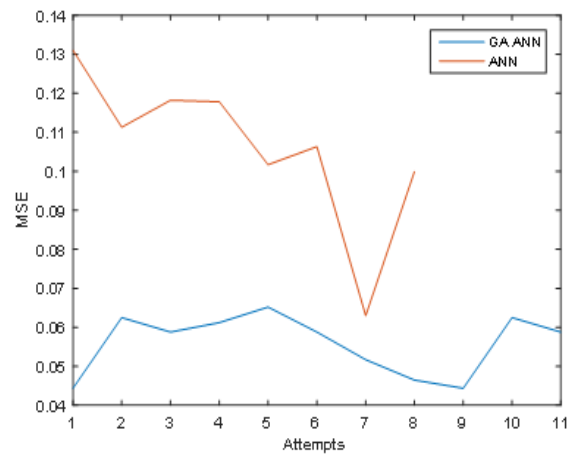


Figure 9. Comparison of mse for ANN and GA-ANN for all attempts

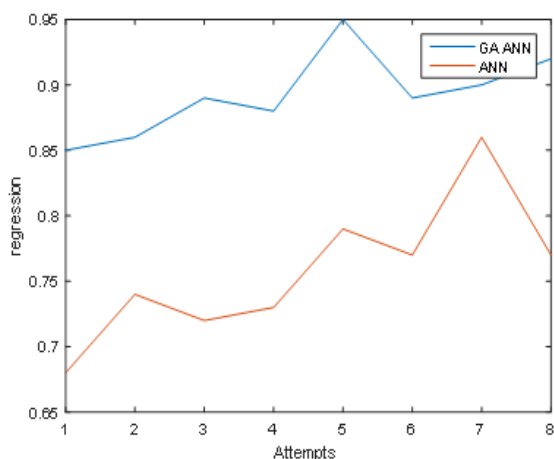


Figure 10. Comparison of regression values for ANN and GA-ANN for all attempts

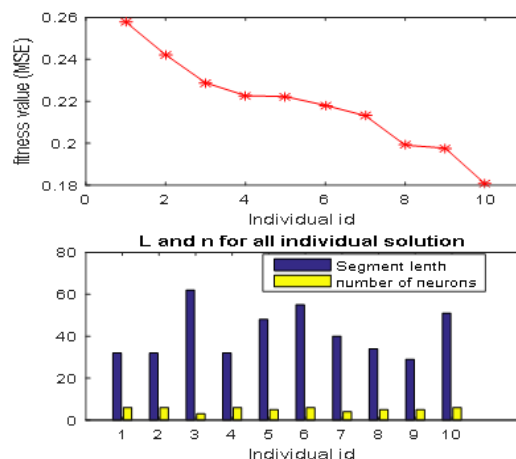


Figure 11. Mse values with respect to individual id

**5. CONCLUSION**

In this work two types of prediction models are developed that can estimate the concentration of carbon mono oxide from the e nose sensor array data. First of all Artificial neural network (ANN) model is developed by manually changing the data segment length and number of hidden nodes. There after the task of ANN design is performed by Genetic algorithm (GA).The perturbation of the segment length and number of nodes is decided as per rule of survival of fittest. The network performance is measured in terms of Mean squared error and regression and it is observed the performance of Genetic algorithm-artificial neural network based data estimator output observes lesser value of mean squared error and higher value of regression than by applying artificial neural network based modeled data estimator. GA-ANN also takes less time and negligible human intervention in setting the parameters in comparison to the ANN model alone. The optimized better performance network model is used to estimate the concentration of Carbon monoxide in a polluted environment.

**ACKNOWLEDGEMENT**

Authors are thankful to Integral Information and Research Centre of Integral University Lucknow, India for providing an opportunity to carry out this research work. This work is an intellectual property of Integral University vides the Manuscript Communication NO. IU/R&D/2019-MCN 000765.

**REFERENCES**

- [1] S. Lee and C. Wu, "Exploring the Vague Environmental Factor via Data-Driven Methods," *2012 8th International Conference on Natural Computation (ICNC 2012)*, pp 409-417, 2012.
- [2] J. Huang, et al., "A Crowdsourc-Based Sensing System for Monitoring Fine-grained Air Quality in Urban Environments," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3240-3247, 201.
- [3] B. Wang, et al., "Air Quality Forecasting Based on Gated Recurrent Long Short Term Memory Model in Internet of Things," *IEEE Access*, vol.7, pp. 69524-69534, 2019.
- [4] M. Jia, et al., "Adopting Internet of Things for the development of smart buildings: A review ofEnabling technologies and applications," *Automationin Construction*, vol. 101, pp. 111-126, 2019.
- [5] R. Montella, et al., "Workflow-based automatic processing for Internet of things crowd sourced data," *Future Generation Computer System*, vol. 94, pp. 103-119, 2019.
- [6] Q. Han, et al., "A Wireless Sensor Network for Monitoring Environmental Quality in the Manufacturing Industry," *IEEE Access*, vol.7, pp. 78108-78119, 2019.
- [7] Ministry Environ. Natural Resour, "Impact Fuel Changes Exhaust Evaporative Emissions," Urban Air Quality Management Project, Sri Lanka, pp. 75-85, 2003.
- [8] WHO, "Chemical Hazards," World Health Organization, Geneva, Switzerland, 2005.
- [9] X. Luo and J. Yang, "Problems and challenges in water pollution monitoring and water pollution source localization using sensor networks," in *2017 ChineseAutomationCongress (CAC)*, Jinan, China, pp. 5834-5838, 2017.
- [10] C. Sun, et al., "Multi-type sensor placements in Gaussian spatial elds for environmental monitoring," *Sensors*, vol. 19, no. 1, p. 189, Jan 2019.

- [11] Q. Han, et al., "Publishing histograms with outliers under data differential privacy," *Security and Communication Networks*, vol. 9, no. 14, pp. 2313-2322, 2016.
- [12] P. Soh, et al., "Adaptive Deep Learning-Based Air Quality prediction Model Using the Most Relevant Spatial-Temporal Relations," *IEEE Access*, vol. 6, pp. 38186-38199, 2018.
- [13] X. Li and A. G. Yeh, "Neural-network-based cellular automata for simulating multiple land use changes using GIS," *International journal of geographical information science*, vol. 16, no. 4, pp. 323-343, 2002.
- [14] G. Jasinski, et al., "Evaluation of the Electronic Nose Used for Monitoring Environmental Pollution," *2018 XV International Scientific Conference on Optoelectronic and Electronic Sensors*, pp. 1-4, 2018.
- [15] J. W. Gardner and P. N. Bartlett, "A brief history of electronic noses," *Sensor and Actuators B: Chemical*, vol. 18, no. 1-3, pp. 210-211, 1994.
- [16] A. Boubrima, et al., "Optimal WSN Deployment Models for Air Pollution Monitoring," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2723-2735, 2017.
- [17] A. Chakma, et al., "Image-based air quality analysis using deep convolutional neural network," *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 3949-3952, 2017.
- [18] K. B. Shaban, et al., "Urban Air Pollution Monitoring System with Forecasting Models," *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2598-2606, 2016.
- [19] U. Gehring, et al., "Traffic-related air pollution and the development of asthma and allergies during the first 8 years of life," *American Journal Respiratory and Critical Care Medicine*, vol. 181, no. 6, pp. 596-603, 2010.
- [20] L. E. Plummer, et al., "Impact of air pollution on lung inflammation and the role of toll-like receptors," *International Journal of Interferon, Cytokine and Mediator Research*, vol. 4, pp. 43-57, 2012.
- [21] S. T. Lee, et al., "Artificial neural network and cellular automata as a modelling simulation for night market spatial development," *The proceeding of 2009 International Association of Societies of Design Research*, 2009.
- [22] A. J. Heppenstall, et al., "Genetic algorithm optimization of an agent-based model for simulating a retail market," *Environment and Planning B: Planning and Design*, vol. 34, no. 6, pp. 1051-1070, 2007.
- [23] M. M. Fischer, "Computational neural networks: a new paradigm for spatial analysis," *Environment and Planning A*, vol. 30, no. 10, pp. 1873-1891, 1998.
- [24] P. W. Oluwasanya, et al., "Portable Multi-Sensor Air Quality Monitoring Platform for Personal Exposure Studies," *IEEE Instrumentation & Measurement Magazine*, vol. 22, no. 5, pp. 36-44, 2019.
- [25] Z. Wang, et al., "Air Quality Measurement Based on Double-Channel Convolutional Neural Network Ensemble Learning," *IEEE Access*, vol. 7, pp. 145067-145081, 2019.
- [26] Y. Zhou, et al., "Data-Driven Air Quality Characterization for Urban Environments: A Case Study," *IEEE Access*, vol. 6, pp. 77996-78006, 2018.
- [27] S. Mahajan, et al., "Improving the Accuracy and Efficiency of PM2.5 Forecast Service Using Cluster-Based Hybrid Neural Network Model," *IEEE Access*, vol. 6, pp. 19193-19204, 2018.
- [28] Y. Wang and G. Chen, "Efficient Data Gathering and Estimation for Metropolitan Air Quality Monitoring by Using Vehicular Sensor Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7234-7248, 2017.
- [29] R. Patuelli, et al., "Neural networks and genetic algorithms as forecasting tools: a case study on German regions," *Environment and Planning B: Planning and Design*, vol. 35, no. 4, pp. 701-722, 2008.
- [30] M. M. Fischer and Y. Leung, "A genetic-algorithms based evolutionary computational neural network for modelling spatial interaction data," *The Annals of Regional Science*, vol. 32, no. 3, pp. 437-458, 1998.
- [31] R. S. Sexton, et al., "Toward global optimization of neural networks: A comparison of the genetic algorithm and back propagation," *Decision Support Systems*, vol. 22, no. 2, pp. 171-185, 1998.
- [32] A. K. Nag and A. Mitra, "Forecasting Daily Foreign Exchange Rates Using Genetically Optimized Neural Networks," *Journal of Forecasting*, vol. 21, no. 7, pp. 501-511, 2002.
- [33] G. P. Zhang, "An investigation of neural networks for linear time-series forecasting," *Computers & Operations Research*, vol. 28, no. 12, pp. 1183-1202, 2001.
- [34] C. M. Bishop, "Neural networks for pattern recognition," Oxford University Press, 1995.
- [35] R. O. Duda, et al., "Pattern Classification," John Wiley & Sons Inc., 2001.
- [36] Lei, et al., "The application of Genetic algorithm toolbox," Xidian University press, Xi'an, 2005.
- [37] S. Openshaw, "Neural network, genetic, and fuzzy logic models of spatial interaction," *Environment and Planning A*, vol. 30, no. 10, pp. 1857-1872, 1998.
- [38] Jou, "The principles and applications of Genetic Algorithm," QuanHun Press, Taipei, 2005.
- [39] H. Schaffers, et al., "Smart cities and the future Internet: Towards cooperation frameworks for open innovation," in *The Future Internet*, pp. 431-446, 2011.

## BIOGRAPHIES OF AUTHORS



**Shirish Pandey** did his B.E. Degree from Shivaji University He obtained M.Tech Degree from N.I.E.L.I.T. He has a vast experience in Digital Design. His strong Areas are Digital Design, Artificial Intelligence Microprocessor and Microcontrollers, VLSI. Presently He is pursuing his Phd. from Integral University Lucknow India





**Dr. S.H.Saeed** received B.E from AMU Aligarh and PhD from Integral University. Presently he is professor and head of Department of ECE, Integral University. He is an excellent record in academics for more than one Decade and also having vast industrial experience. He has published many books, research papers in reputed journals at national and international level.



**Naimur Rahman Kidwai** received the B.Sc.(Engg.) degree in electronics engineering from the ZakirHussain College of Engineering and Technology, Aligarh Muslim University, India, in 1996, the M.Tech. Degree in digital communication from Uttar Pradesh Technical University, Lucknow, India, in 2006, and the Ph.D. degree from Integral University, Lucknow, in 2014. He is currently an Associate Professor with the Department of Electronics and Communication, Integral University.