BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **23**, No 4
Sofia • 2023 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2023-0046

Reliability Analysis of an IoT-Based Air Pollution Monitoring System Using Machine Learning Algorithm-BDBN

Saritha, Sarasvathi V.

Department of Computer Science and Engineering, PESIT-Bangalore South Campus, Bangalore and Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India
E-mails: saritha.k@pes.edu sarsvathiv@pes.edu

Abstract**:** *Transmission of information is an essential component in an IoT device for sending, receiving, and collecting data. The Smart devices in IoT architecture are designed as physical devices linked with computing resources that can connect and communicate with another smart device through any medium and protocol. Communication among various smart devices is a challenging task to exchange information and to guarantee the information reaches the destination entirely in real-time in the same order as sent without any data loss. Thus, this article proposes the novel Bat-based Deep Belief Neural framework* (*BDBN*) *method for the air pollution monitoring scheme. The reliability of the proposed system has been tested under the error condition in the transport layer and is validated with the conventional methods in terms of Accuracy, Mean Absolute Error* (*MAE*)*, Root Mean Square Error* (*RMSE*)*, Pearson correlation coefficient* (*r*)*, Coefficient of determination* (R^2) *and Error rate.*

Keywords**:** *Air quality monitoring, Metaheuristic optimization, Machine learning, Internet of Things, Pollution monitoring scheme, Cloud storage.*

1. Introduction

Air Quality (AQ) is a necessary concern among people due to its impact on the environment and human beings. Also, air pollution creates many diseases, one of the foremost reasons for death. Thus, an Air Quality Monitoring (AQM) system is necessary to reduce the adverse properties of air pollutants [1-3]. Also, the calculation and tracking of AQM are continuously employed to minimize air pollution's effects. Moreover, the AQM scheme is based on indoor and outdoor monitoring, which denotes the location or area [4]. Indoor tracking depends on the air pollution inside offices, homes, workplaces, etc. [5]. Outdoor monitoring depends on the air pollution in the open places, environmental surroundings, or the location of industries [6]. To compare indoor and outdoor tracking, the outdoor AQM is very complex [7]. Moreover, the Air Quality Index (AQI) calculation is employed to identify the original state of pollution [8].

Nowadays, cloud computing and the Internet of Things (IoT) can monitor real-time processes in different sectors. Moreover, it is utilized for AQM in the indoor environment, analyzing and visualizing the data [9-10]. Additionally, consistent analytical data about the air quality has been observed for using pollutant variables. The important pollutants to reduce the AQ are Carbon Monoxide (CO), Particulate Matter (PM), ground-level Ozone (O_3), and Sulphur Dioxide (SO_2) [11]. In this, the pollutants are identified by the communication protocols for short-range like ZigBee, Bluetooth, etc., and the development of IoT utilizes the gas sensors for a wide area [12]. Thus, the sensors are utilized for measuring the levels of temperature or humidity and gases like CO, SO_2, and PM [13-14].

Machine Learning (ML) and optimization-based methods are mainly part of the monitoring system. The methods previously used include Azure ML dynamic Fuzzy, Multi-objective Particle Swarm Optimization (MPSO) with the Adaptive Neuro-Fuzzy combined Inference Scheme (ANFIS), multi-objective optimization Distributed Intelligence [21-25], etc. These methods predict environmental conditions like temperature, gases, and PM. If a specific value increases the pollutant level of air, it creates an alert to the user. However, these existing approaches have some limitations like low accuracy rate, low range monitoring, and high-cost sensors [15].

This article proposes a novel cost-significant air quality monitoring and risk factor prediction model using the BDBN method. The work used optimization fitness function combined with the ML approach, IoT sensor-based AQ data. The proposed model's reliability is validated under disturbances and is compared with the various existing methods.

The article is structured as follows: the recent literature related to this article is explained in Section 2. In Section 3, the system model, along with its problem statement, is described. The proposed method and its workflow are elaborated in Section 4. In Section 5, the result, as well as a discussion of this work, is detailed. Then, the paper is concluded in Section 6.

2. Related work

Some of the recent works based on air pollution monitoring systems using IoT are detailed below.

Air pollution is one of the important risks to human health as well as the environment. Sharafat Ali et al. [16] have developed a novel sensor node with low-cost electrochemical sensors and infrared sensors for measuring Nitrogen Dioxide (NO_2), CO concentrations, and Particulate Matter (PM) levels. The sensor node being developed utilizes the power from the main supply or solar, which is applicable for long-range and short-range communication. However, this model could not identify the interference in the air quality.

Subsequently, Senthilkumar, Venkatakrishnan, and Balaji [17] have introduced a distinct model using IoT-based fog computing for monitoring air quality. This model also utilized the embedded model for collecting data about air quality at a particular time by sensors that are sent to the fog nodes. However, this fog computing-based IoT model is unsuitable for measuring air quality in an extensive range.

Additionally, a distributed intelligence-based AQM model has been developed by Lazrak et al. [18], which measures the air quality in urban areas. This model uses node interoperability, collective learning, and shared knowledge to attain better results in monitoring and prediction. This approach cannot predict in any way the various categories of gases like CO, NO_2, and PM.

The existing AQM models have limitations like low fabrication repeatability, interference of unexpected gas, and sensor poisoning. Saverio De Vito et al. [19] have introduced an adaptive machine learning model in the non-stationary structure by multi-sensors calibration scheme to overcome these issues. Thus, the approach achieved high-resolution air quality mapping in urban areas. In this, a backup system has been used for attaining data when the AQM data is unreachable due to maintenance and faults. However, this model cannot identify the interruption of noisy updates.

Moreover, Purkayastha et al. [20] have introduced an AQM model by developing a cloud server to store air quality information. Here, the data about the air quality are measured from different categories of data sources with the use of sensors that are integrated by microcontrollers. This model's process is complicated and takes significant time.

3. Research problem

AQ is one of the most important concerns for people because it affects the environment and humans. The basic AQ prediction process system model is detailed in Fig. 1. Several AQM schemes have been developed to monitor air quality, but these methods use high-cost sensors. Moreover, many approaches have been developed with low-cost sensors to monitor AQ, but they have some limitations like low accuracy, short-range pollution measurement, etc.

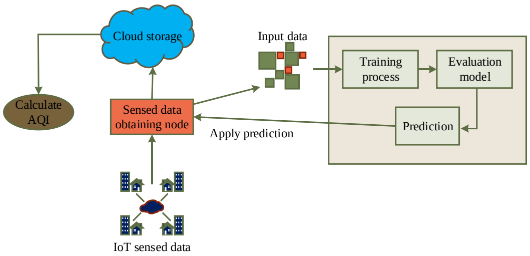


Fig. 1. System model of AQ prediction

4. Methodology

Air pollution is one of the leading reasons for the death of human beings worldwide, and the existing AQM schemes have some limitations like high cost and inaccurate prediction, which create harmful effects. Thus, this research introduces a novel BDBN to improve the AQM system's efficiency and achieve reliable analytical information on the AQ. Here, the fitness function of the bat optimization is updated in the DBN framework for attaining efficient output for AQM. The IoT sensors have been also utilized for identifying the rate of AQI for pollutants like CO, NO_2, and PM using the proposed BDBN model. The gathered data is kept in a cloud database, offering scalability, security, and accessibility. It also monitors and analyzes air pollution data in the pre-processing phase to remove the noise and error values from the collected data. Subsequently, a novel BDBN model is designed to validate and predict the air pollution rate. Malicious activities are launched in the transport layer of the BDBN model to test the reliability of the proposed model. Finally, the results are calculated to prove the proposed BDBN model's reliability and the proposed method's process is detailed in Fig. 2.

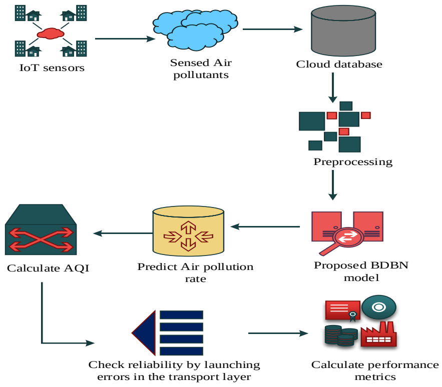


Fig. 2. Process of the proposed methodology

4.1. Data collection

Usually, the IoT-based sensors are connected to the cities. IoT sensors sense the air quality and data is stored in the cloud database. Here, AQ dataset of India is used for carrying the performance of Machine learning model.

4.2. Pre-processing

The gathered data from the AQ database may be enclosed with the missing values and repeated and inconsistent data. Thus, the dataset should be cleared for accurate AQ prediction outcomes by deleting or providing the mean values in the place of missing rates. The elimination of inconsistent data avoids the unfairness of the consequences. In pre-processing, the extreme or outlier values are eliminated for optimal prediction. After pre-processing, machine learning will work for accurate forecasts.

4.3. Proposed BDBN model

In monitoring AQ, the choice of BDBN model monitoring is significant due to its influential factors such as the rapid convergence at a very primary phase, high accuracy, and the solution to the gradient issues. Initially, the dataset is divided into testing and training groups. The predicting AQ model is first trained with the training dataset and then it is validated with the test dataset. Afterward, the tested data performance is estimated by various parameters. The BDBN model combines bat optimization and a deep belief-based system. This model has diverse layers such as an input layer, different hidden layers, and an output layer. This algorithm is enclosed with the visible layer's input and output sides and hidden points at the hidden layers. Consider V is denoted as the Visible node and H is denoted as the Hidden node, respectively. The V node's values are in binary. The Probabilities for the Visible (PV) are estimated using the equation

$$\text{PV} = \frac{1}{N}\sum_{H} e^{-S(V,H)}, \tag{1}$$

where: N is denoted as the normalizing function; $S(V, H)$ is denoted as the energy portion of the visible as well as hidden section. "Energy portion of the visible in the hidden section" refers to the term $S(V, H)$, which represents the contribution of the visible units to the energy function when the hidden units are activated. This term measures how well the visible units align with the connections (weights) to the hidden units. Then, the weight of the AQ data is provided to the training sector, which is performed by the gradient descent method using the next equation:

$$W(n+1) = W(n) + \alpha \frac{\partial \log(\text{PV})}{\partial W}, \tag{2}$$

where $\frac{\partial \log(\text{PV})}{\partial W}$ is the gradient portion that is estimated by the average separation of trained data and tested data, n is denoted as the number of iterations, α is the learning rate [27]. The Probabilities for the Hidden layer (PH) are estimated based on the parallel data using the equation

$$\text{PH} = P(H = 1 | V) = \beta(u + \sum WV), \tag{3}$$

where the logistic sigmoid action is denoted as β, W is the weight rate, u is denoted as a bias set for each node. Consequently, the state of binary is selected for the hidden portion, and then the condition of the visible unit is one, and that probability for the visible layer is estimated based on the parallel data using the equation

$$\text{PV} = P(V = 1 | H) = \beta(u + \sum WV). \tag{4}$$

Then the training function is started from the visible layer data. Afterward, the sample data in the hidden layer is conducted using Equation (3). This process is continuous, and the data of visible and hidden are stored and form a new state V' and H' which is considered as in the next equation,

$$\Delta W = \alpha\left(\langle VH \rangle - \langle V'H' \rangle\right), \tag{5}$$

α is denoted as the rate of learning, and the training data expectation is denoted as $\langle . \rangle$. The data is split into trained and test data. The test data is predicted AQ data. For the accurate prediction state, the data weights are updated using

$$\Delta W = \langle V'H' \rangle + \alpha \left(\langle VH \rangle_{\text{data}} - \langle V'H' \rangle \right)_{\text{reconstruction}} . \tag{6}$$

Then, robust and rapid validation is essential due to the impact of day-by-day environmental changes. Thus, the speed of AQI prediction is estimated using the equation

$$S^t = \left[S^{t-1} \right] \times (A^t - \hat{A}) \Delta W , \tag{7}$$

where S is denoted as the speed, and t is denoted as time. Furthermore, the fitness function of the bat is established to the deep belief approach for accurate AQ prediction. The exact prediction of AQ is estimated using the equation

$$A^t = \beta \times y + a , \tag{8}$$

a represents the predicted AQI coefficient constant, A^t presents the predicted AQI coefficient, β represents the slope parameter, y represents the quantity of a particular pollutant [28].

4.4. Estimate AQI for risk factor

The value of AQI is used to estimate the quantity of air pollution and risk factors. Each place has its quality index value. In this research, the European standard of AQI is considered. Moreover, the sub-index of AQ is estimated based on the average pollution range in day-to-day life and the average concentration of each place daily. The CO, NO_2, O_3, and PM rates have been re-estimated in all places. The Risk Factor (RF*) is calculated using the next equation by the summation of D_r^t,

$$\text{RF}^* = \sum D_r^t , \tag{9}$$

$$D_r^t = T_t^* \frac{B_t}{B_o} , \tag{10}$$

where: T_t^* reflects the risk of pollution of air in the atmosphere; B_t denotes the average pollution in the day-to-day life; B_o reflects the average concentration of each place daily. If the above condition is not satisfied, then the new AQ monitoring is created using the equation

$$A_{\text{anew}}^t = A_{\text{aold}}^t + rQ^t . \tag{11}$$

Algorithm 1. Proposed BDBN in AQ monitoring system

Input: Data of CO, NO_2, and PM *// sensed by IoT sensors*
Pre-processing the AQ data *// removing the unwanted data*
Initialize all the data parameters to the BDBN method
Consider V and H *// visible and hidden node*
Estimate the probabilities for the visible
// energy portion of visible as well as hidden section and normalization
Apply gradient descent method *//training the data*
Estimate probabilities for the hidden layer
Provide the condition for the visible data
The weights of visible and hidden data stored by Equation (5)
Update the state of the weights of the data

Apply AQI prediction speed
Exact AQI prediction using Equation (8)
Calculate the risk factor
If
$\mathrm{RF}^{*} = \sum D_r^t$
Solution attained
Else
$A_{\text{anew}}^t = A_{\text{aold}}^t + rQ^t$
Output: Optimal AQ monitoring is achieved

5. Result and discussion

The simulation of this proposed model is done in MATLAB on the Windows platform. The parameters used for the AQM in BDBN method are illustrated in Table 1.

Table 1. Parameters used for the AQM in BDBN method

Parameter	Values
Population size	1000
Number of iterations	100
Pulse rate	0.4
Visible node	4
Hidden node	5

5.1. Case study

Assume A is the urban region in which the IoT-based sensors have been placed at different regions. The sensors have sensed the data of AQ, such as the CO, NO_2, O_3, and PM rates. The collected data has been stored in cloud storage. This study consumes the IoT-based AQ sensed data from the UCI dataset. The dataset in the file is arranged in the order of hourly measured information, separated by different parameters and pollutant data such as CO, NO_2, O_3, and PM observation from the region of India.

The overall hourly data has been gathered from 2015 to 2020 and 30,000 data were used. Then, the collected data was pre-processed by eliminating the unwanted data and filling the missing data based on the average rate among the earlier and the coming day records. In the missing data, the place has been filled with the value zero.

Moreover, the dataset column has been added to classify whether AQ is a pollutant or not and predict the risk factor. From the overall data set, 70% of data is taken for the training phase, and 30% of data is considered for the testing/validation stage. The data is provided to the proposed BDBN algorithm for monitoring the AQ. The category of AQI and health risk factor details is demonstrated in Table 2. The flowchart of the proposed AQI monitoring system is illustrated in Fig. 3. The performance metrics are estimated for the validation of the proposed system approach.

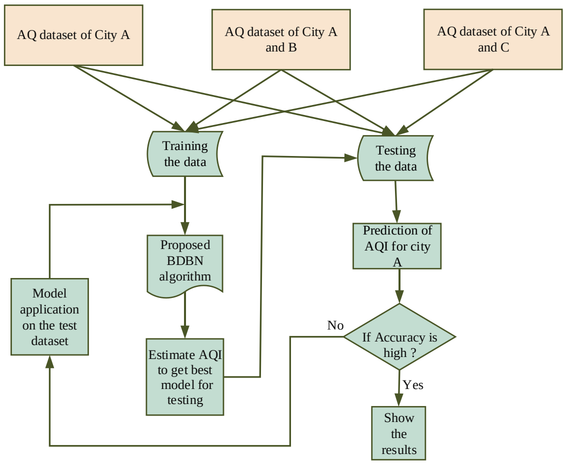


Fig. 3. Flowchart of proposed AQI monitoring system

Table 2. Category of AQI and health risk factor

Category (Range)	CO (8 h)	NO_2 (24 h)	PM (24 h)	O_3 (8 h)	Health risk factor
Good (0-50)	0-1.0	0-40	0-50	0-50	Less impact
Reasonable (51-200)	1.1-2.0	41-80	51-100	51-100	Create minor breathing
Medium (201-300)	2.1-10	81-180	101-250	101-168	Discomfort with Asthma and heart disease people
Bad (201-300)	10-17	181-280	251-350	169-208	Prolonged inhaling and heart disease people are highly affected
Very bad (301-400)	17-34	281-400	351-430	209-748	Create respiratory problems, affect lungs, and highly suffered by heart diseases
Severe (401-500)	>34	>400	>430	>748	Create respiratory problems, affect the lungs, and suffer from heart disease even a healthy person

5.2. Performance analysis

The proposed IoT-based AQ monitoring by the BDBN framework is validated with different parameters. The performance metrics are estimated to prove the proposed model's efficiency, accuracy, MAE, RMSE, R^2, and error rate.

5.2.1. Accuracy

The entire amount of actual AQ prediction rate is estimated by accuracy rate, which is estimated as,

$$\text{Accuracy} = \frac{\text{TruePositive} + \text{TrueNegative}}{\text{TruePositive} + \text{TrueNegative} + \text{FalsePositive} + \text{FalseNegative}}. \tag{12}$$

5.2.2. Mean Absolute Error (MAE)

The MAE is the estimated number of errors among the balancing statements expressing the identical phenomenon, which is estimated using

$$M_{\mathrm{e}} = \frac{\sum_{i=1}^{n} |p_i - t_i|}{n}, \tag{13}$$

where p_i is the predicted air quality and t_i is the true rate of air quality [29].

5.2.3. Pearson correlation coefficient (r)

The Pearson correlation coefficient r is utilized to estimate the association quality among two data variables. The covariance of the two variables is split by the product of their respective standard deviations, and it is used to measure the linear dependence between the two variables, which is estimated as,

(14) $$r = \frac{\mathrm{cov}(P_\mathrm{c}, O_\mathrm{r})}{\sigma_{P_\mathrm{c}}, \sigma_{O_\mathrm{r}}},$$

where the concentration of the predicted pollutant is represented as P_c, the observed rate is denoted as O_r, the standard deviation is represented as σ and the covariance is denoted as cov. In the analysis of r, the values should be within the range of [–1, 1]; however, if the value is zero, then it shows as no connection.

5.2.4. RMSE

The RMSE is evaluated to analyze the square root of the average square error among the observation and predicted air quality values that are calculated using the equation

(15) $$R_m = \sqrt{\left(\frac{1}{N}\right)\sum_{i=1}^{N}\left(p_i - t_i\right)^2},$$

where, the squared error among the observed as well as the predicted rate at the data point i is denoted as $\left(p_i - t_i\right)^2$ and N is the overall amount of data in the system [30].

5.2.5. Correlation coefficient

The correlation coefficient R^2 is utilized to estimate in what way the an achieved data variable is the difference by the observed variable variation. The rate of R^2 is calculated using the equation

(16) $$R^2 = 1 - \frac{R_\mathrm{s}}{T_\mathrm{s}}.$$

Square of residual summation is represented as R_s and the overall sum of squares is denoted as T_s [31].

5.2.6. Error rate

The error rate is estimated as the ratio of the amount of incorrect AQ predictions to the overall amount of the dataset. The finest error is considered as less than 1%.

5.2.7. Reliability estimation

The proposed model analysis and properties of individual predictions are defined by the reliability estimation, which is expressed using

(17) $$\text{Reliability} = 1 - \text{Error rate}.$$

Initially, the proposed DBDN method is applied for the AQ monitoring. Then, the parameters of metrics are demonstrated in Fig. 4.

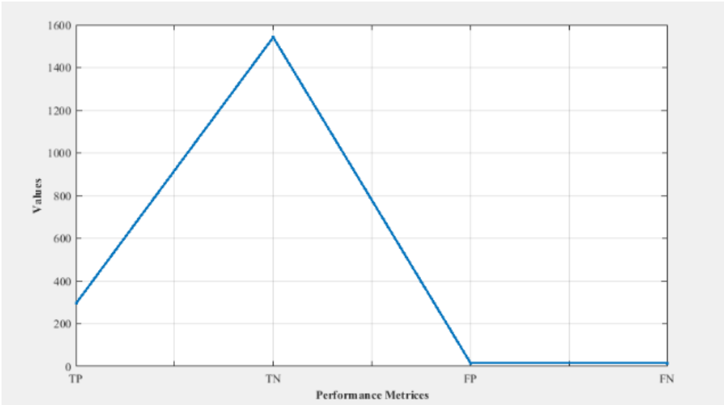
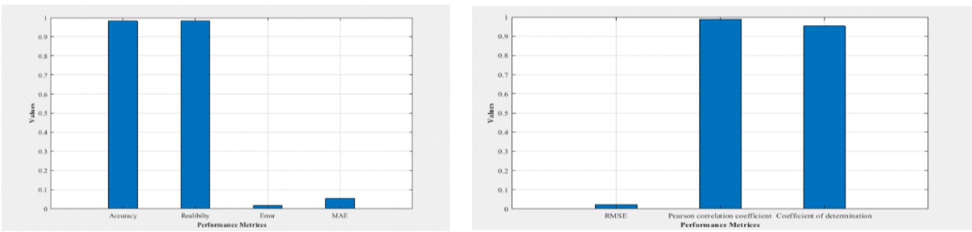


Fig. 4. Performance parameter metrics

Consequently, the performance metrics of accuracy, MAE, RMSE, R^2, Pearson correlation coefficient of determination, and error rate are estimated which is demonstrated in Fig. 5. The accuracy of the proposed BDBN method is attained as 0.9817, and the MAE value is 0.0534. The attained RMSE value is 0.23125, and the reduced error rate is 0.0178. Moreover, the improved value of R^2 and Coefficient of determination are achieved as 0.98823 and 0.95346 rates. The reliability of the estimated proposed system is obtained as 0.98217.

Then, the proposed BDBN is removed, and the performance of AQ monitoring is validated with the usual deep belief method. Consequently, the performance of the system without the proposed BDBN method is illustrated in Fig. 6.

Consequently, the performance metrics of accuracy, MAE, RMSE, R^2, Pearson correlation coefficient of determination, and error rate have been estimated without the BDBN method, which is demonstrated in Fig. 7. The accuracy of the proposed BDBN method is attained as 0.95152, and the MAE value is 0.14545. The attained RMSE value is 0.038139, and the reduced error rate is 0.048485. Moreover, the improved value of R^2 and coefficient of determination are achieved at 0.96798 and 0.876 rates.



(a) (b)

Fig. 5. Performance metrics evaluation of the proposed approach

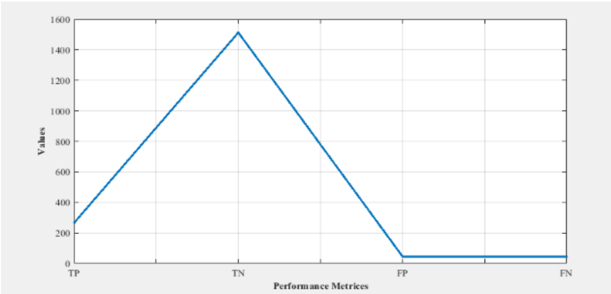
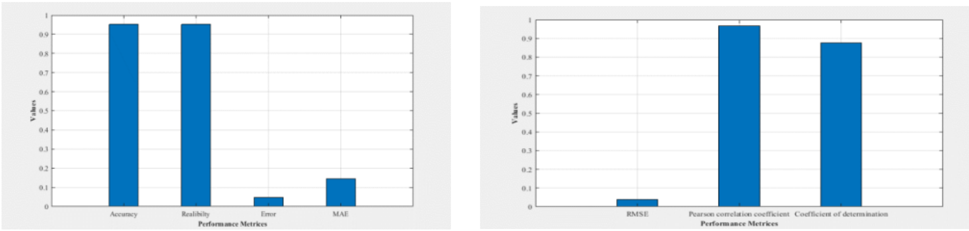


Fig. 6. Performance metrics without BDBN method



(a) (b)

Fig. 7. Performance metrics evaluation without BDBN method

With and without the proposed BDBN comparative analysis is considered for validating the effectiveness of the proposed hybrid system. Consequently, the performance of the system with and without proposed BDBN method is illustrated in Fig. 8 and Table 3.

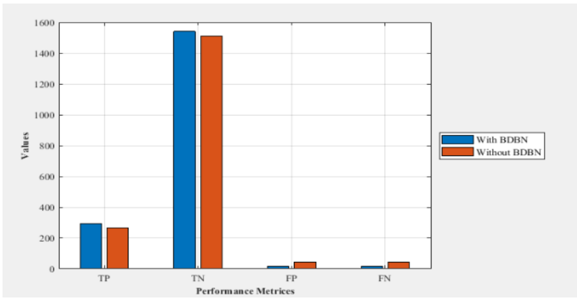
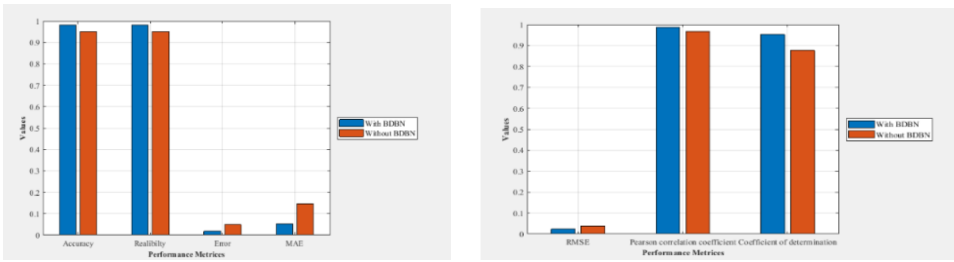


Fig. 8. Performance metrics evaluation for without BDBN method

Table 3. The comparative analysis of performance metrics for with and without BDBN method

Methods and metrics	TP	TN	FP	FN
With BDBN	295	1541.667	16.66667	16.66667
Without BDBN	266.3333	1513	45.33333	45.33333

Consequently, the performance metrics of accuracy, MAE, RMSE, R^2, Pearson correlation coefficient of determination, and error rate are compared for with and without BDBN method, and that is demonstrated in Fig. 9. The accuracy of the proposed BDBN method is attained as 0.98217, and without BDBN method is 0.95152 and the MAE value for the proposed system is 0.053476 and without BDBN method is 0.14545. The attained RMSE value is 0.023125 for the proposed BDBN and without BDBN method is 0.038139 and the reduced error rate for the proposed system is 0.017825 and without BDBN method is 0.048485. Moreover, the improved value of R^2 and Coefficient of determination are achieved with and without BDBN method at 0.98823 and 0.953 as well as 0.96798 and 0.876 rates.



(a) (b)

Fig. 9. Performance metrics comparison with and without BDBN method

5.2.8. Comparative analysis

Also, the proposed BDBN method is compared with the different existing methods such as DI(RF) [11], ANN [16], and Reinforcement learning and Genetic algorithm (RL-GA) [26]. Accuracy comparison with existing methods is shown in Fig. 10. Conventional methods like DI(RF) have attained 92% of accuracy, ANN is 89% and PL-GA is 90% which is less over the proposed method. Without BDBN method, the accuracy is also estimated and obtained as 95.152%. However, the conventional method have improved the accuracy as 98.17%. This shows the effective monitoring of AQ system.

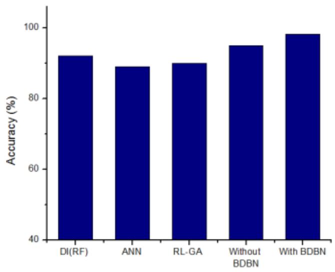


Fig. 10. Accuracy comparison with existing methods

The MAE of proposed method is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] are exposed in Fig. 11. The proposed BDBN method has achieved very less 0.0534 MAE value, which is less over the existing methods. The existing DI(RF) [11] method has attained 4.197 of MAE, which is very high over other ANN [16] and RL-GA [26] methods. The ANN method has achieved 3.7 MAE and 0.06 for the RL-GA technique. Without BDBN method, the MAE is also estimated and has obtained as 0.14545.

The correlation coefficient of the proposed method is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] revealed in Fig. 12. Here, the comparison demonstrates that the proposed BDBN method has attained a very high 0.95346 of *r*-value. The existing DI(RF) method has a 0.95 of *r*-value and 0.81 for the ANN method. The RL-GA method has achieved a 0.934 of *r*-value. Also, without BDBN method, the *r*-value is estimated and obtained as 0.96798. Nevertheless, the proposed system has attained better performance.

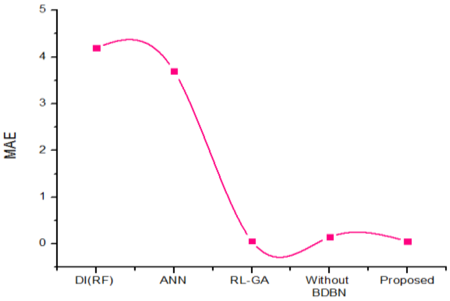


Fig. 11. MAE comparison with existing methods

The RMSE of the proposed method is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] shown in Fig. 13. The conventional DI(RF) has attained 7.775 RMSE that is very high over other methods. The ANN and RL-GA method has attained 0.075 and 4.32 of RMSE, but the proposed method has achieved much less than 0.23125 of RMSE. Also, without the BDBN method, the *r*-value is estimated and obtained as 0.038139. The RMSE from the projected system is much less; thus, it shows the effective AQ estimation range.

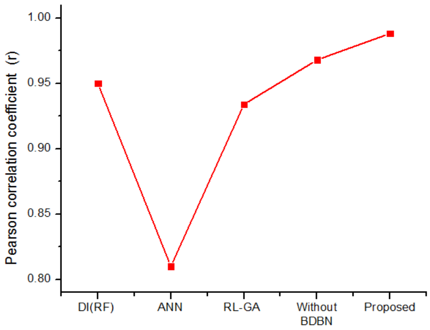


Fig. 12. Pearson correlation coefficient comparison with existing methods

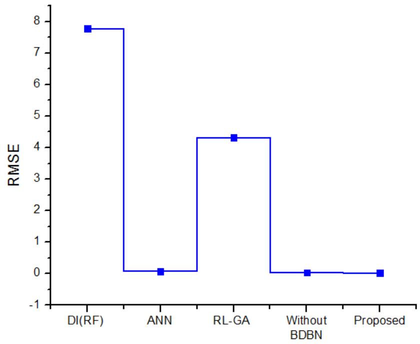


Fig. 13. RMSE comparison with existing methods

The correlation coefficient of the proposed method is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] are shown in Fig. 14. The existing DI(RF) [11] method has achieved 0.953 of R^2, the traditional ANN [16] has obtained 0.86 of R^2, and RL-GA has 0.92 of R^2 is got. Without BDBN method, the value of R^2 is also estimated and obtained as 0.876. Also, without the BDBN method, the value of R^2 is estimated and obtained as 0.876. However, the proposed BDBN method has achieved 0.98823 of R^2. This is a very high rate over the conventional methods.

The error rate obtained from the proposed method is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] shown in Fig. 15. The error rate of the DI (RF) method is 0.16, ANN method is 0.095, and RL-GA method has attained as in Section 4.6. The proposed DBDN method has obtained a much lower 0.0178 error rate than the other methods. Also, without the BDBN method, the error is estimated and obtained 0.048485. Here, the conventional RL-GA method has attained more error value over the compared techniques.

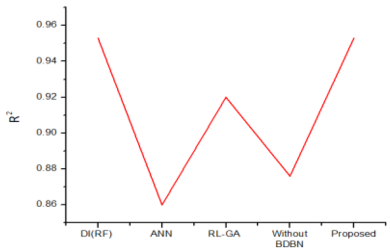


Fig. 14. correlation coefficient comparison with existing methods

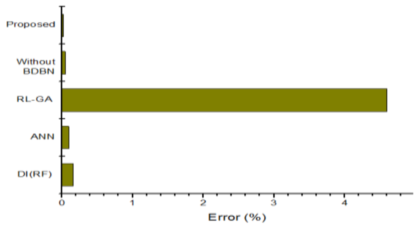


Fig. 15. Comparison of error rate with existing methods

Furthermore, the reliability of the system is compared with the conventional methods like DI(RF) [11], ANN [16], and RL-GA [26] are illustrated in Fig. 16. It shows that the proposed system has attained more reliability than the earlier methods because of the higher reliability rate of 98.217%, which is higher than other methods. Then, the proposed BDBN method is removed only nominal deep belief is applied to validate the performance of the system, but it achieved only 95.12% reliability, which is less than the proposed BDBN technique. Thus, it shows the effective performance of the proposed system.

The overall performance of the proposed system is compared with the conventional methods such as DI(RF) [11], ANN [16], and RL-GA [26] detailed in Table 4. Hence, this comparison shows that the proposed method consumes 98.17% accuracy over other methods; the conventional methods have attained much less accuracy than the proposed BDBN method.

The MAE and RMSE values are much less than the existing DI(RF) [11], ANN [16], and RL-GA [26] methods. Consequently, the value of R^2 and coefficient of determination r are improved than the existing techniques. The elapsed time for the overall computation is 0.417 s. This shows the effective performance of the proposed AQ monitoring strategy.

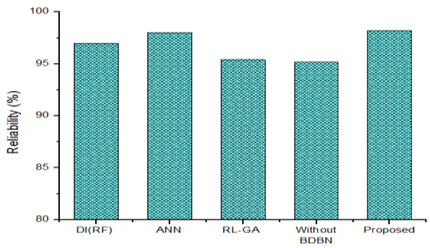


Fig. 16. Comparison of reliability analysis with existing methods

Table 4. Overall performance comparison of the proposed method

Parameters	DI(RF) [11]	ANN [16]	RL-GA [26]	Without BDBN	Proposed
Accuracy (%)	92	89	90	95.152	98.17
MAE	4.197	3.7	0.06	0.14545	0.0534
RMSE	7.775	0.075	4.32	0.038139	0.023125
R	0.95	0.81	0.934	0.96798	0.98823
R^2	0.953	0.86	0.92	0.876	0.953
Error rate	0.16	0.095	4.6	0.048485	0.0178
Reliability (%)	97	98	95.4	95.152	98.217

5.3. Discussion

The discussion demonstrates the merits and limitations of state-of-the-art methods in Table 5. The observation shows that the conventional methods have certain limitations over the proposed BDBN method. The conventional ANN [16] method has consumed more time than the other methods, and the DI(RF) [11] method has attained high skill, but the computational problem of this method is more noticeable. Moreover, the RL-GA [26] has attained a higher error rate than other methods. The reliability of the proposed system is very high compared with the earlier prediction and monitoring strategies.

Table 5. Comparison of state-of-the-art methods

Author	Method	Merits	Limitation
[11]	DI(RF)	Supreme knowledgeable function is achieved	High computational burden
[16]	ANN	Data loss does not affect the performance of the system	Time duration is longer, and a high error rate is attained
[26]	RL-GA	Improved the performance and attained variation for more time interval	Not executing the inner data and required more amount of data
Proposed	BDBN	Reliable performance is achieved even the disturbances occur	Need to improve the accuracy further

Thus, the overall comparative analysis demonstrates that the proposed BDBN method has achieved the finest performance in AQ monitoring system when compared with the conventional methods. It proves also that the proposed BDBN method has consumed more accuracy, fewer error values, and improved coefficient rates. Therefore, the results show the effective performance of the proposed BDBN method in AQ monitoring system.

6. Conclusion

AQ monitoring system is a major concern all over the world due to the rapid growth of industrialization and urbanization. If the level of AQ is not monitored properly, then the life of the people and other organisms is critical and then it leads to crucial diseases and death. Various IoT-based methods have been developed, yet it is not sufficient for accurate estimation. For this reason, a novel IoT-based BDBN method has been developed in this article for monitoring and analyzing the risk factors of AQ. The reliability of the proposed system has been also validated under certain error conditions. The simulation outcomes are compared with the conventional methods in terms of different performance parameters. The accuracy of the proposed method has attained 98.17%, which is very high over existing methods. The reliability of the AQ system is estimated under the conditions with and without the BDBN method. It has obtained better reliability at the proposed BDBN method at 98.217% than the without BDBN approach. The error rate, MSE, and RMSE have also been achieved much less. In the future, hybrid soft computing methods can take place with a high amount of data for the monitoring system under climatic variations.

References

1. Collier-Oxandale, A., et al. Field and Laboratory Performance Evaluations of 28 Gas-Phase Air Quality Sensors by the AQ-SPEC Program. – Atmospheric Environment, Vol. **220**, 2020, 117092.
2. Kassandros, T., et al. Citizens in the Loop for Air Quality Monitoring in Thessaloniki, Greece. Advances and New Trends in Environmental Informatics. Cham, Springer, 2021, pp. 121-130.
3. Chen, J., et al. Modeling Air Quality in the San Joaquin Valley of California during the 2013 Discover-AQ Field Campaign. – Atmospheric Environment. Vol. **X**, 2020, No 5, 100067.
4. Huang, Y., et al. Evaluating In-Use Vehicle Emissions Using Air Quality Monitoring Stations and On-Road Remote Sensing Systems. – Science of the Total Environment, Vol. **740**, 2020, 139868.
5. Srivastava, M., R. Kumar. Smart Environmental Monitoring Based on IoT: Architecture, Issues, and Challenges. Advances in Computational Intelligence and Communication Technology. Singapore, Springer, 2021, pp. 349-358.
6. Pereira, W. F., et al. Environmental Monitoring in a Poultry Farm Using an Instrument Developed with the Internet of Things Concept. – Computers and Electronics in Agriculture, Vol. **170**, 2020, 105257.
7. Salam, A. Internet of Things for Environmental Sustainability and Climate Change. – Internet of Things for Sustainable Community Development. Cham, Springer, 2020. pp. 33-69.
8. Ha, Q. P., S. Metia, M. D. Phung. Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring. – IEEE Sensors Journal, Vol. **20**, 2020, No 8, pp. 4430-4441.

9. G o m e s, J. B., et al. A Novel Internet of Things-Based Plug-and-Play Multigas Sensor for Environmental Monitoring. – Transactions on Emerging Telecommunications Technologies, 2020, e3967.
10. M o t l a g h, N. H., et al. Toward Massive Scale Air Quality Monitoring. – IEEE Communications Magazine, Vol. **58**, 2020, No 2, pp. 54-59.
11. E s f a h a n i, S., et al. Smart City Battery Operated IoT Based Indoor Air Quality Monitoring System. – 2020 IEEE Sensors. IEEE, 2020.
12. M a r q u e s, G., C. R. F e r r e i r a, R. P i t a r m a. Indoor Air Quality Assessment Using a CO2 Monitoring System Based on Internet of Things. – Journal of Medical Systems, Vol. **43**, 2019, No 3, pp. 1-10.
13. D h i n g r a, S., et al. Internet of Things Mobile–Air Pollution Monitoring System (IoT-Mobair). – IEEE Internet of Things Journal, Vol. **6**, 2019, No 3, pp. 5577-5584.
14. K i m, S. H., et al. Development of an IoT-Based Atmospheric Environment Monitoring System. – In: Proc. of International Conference on Information and Communication Technology Convergence (ICTC'2017), IEEE, 2017.
15. D e l a B a r r e r a, F., et al. Megafires in Chile 2017: Monitoring Multiscale Environmental Impacts of Burned Ecosystems. – Science of the Total Environment, Vol. **637**, 2018, pp. 1526-1536.
16. A l i, S., et al. Low-Cost Sensor with IoT LoRaWAN Connectivity and Machine Learning-Based Calibration for Air Pollution Monitoring. – IEEE Transactions on Instrumentation and Measurement, Vol. **70**, 2020, pp. 1-11.
17. S e n t h i l k u m a r, R., P. V e n k a t a k r i s h n a n, N. B a l a j i. Intelligent Based Novel Embedded System Based IoT Enabled Air Pollution Monitoring System. – Microprocessors and Microsystems, Vol. **77**, 2020, 103172.
18. L a z r a k, N., et al. Enabling Distributed Intelligence in Internet of Things: An Air Quality Monitoring Use Case. – Personal and Ubiquitous Computing, 2020, pp. 1-11.
19. D e V i t o, S., et al. Adaptive Machine Learning Strategies for Network Calibration of IoT Smart Air Quality Monitoring Devices. – Pattern Recognition Letters, Vol. **136**, 2020, pp. 264-271.
20. P u r k a y a s t h a, K. D., et al. IoT Based Design of Air Quality Monitoring System Web Server for Android Platform. – Wireless Personal Communications, 2021, pp. 1-20.
21. W a n, J. B., R. M. A. K h a n. An Internet of Things System for Underground Mine Air Quality Pollutant Prediction Based on Azure Machine Learning. – Sensors, Vol. **18**, 2018, No 4, 930.
22. X u, Y., P. D u, J. W a n g. Research and Application of a Hybrid Model Based on Dynamic Fuzzy Synthetic Evaluation for Establishing Air Quality Forecasting and Early Warning System: A Case Study in China. – Environmental Pollution, Vol. **223**, 2017, pp. 435-448.
23. S i n g h, P. H., D. S i n g h, A. K. M a l h i. Multi-Objective Particle Swarm Optimization-Based Adaptive Neuro-Fuzzy Inference System for Benzene Monitoring. – Neural Computing and Applications, Vol. **31**, 2019, No 7, pp. 2195-2205.
24. B a r o t, V., V. K a p a d i a, S. P a n d y a. QoS Enabled IoT Based Low-Cost Air Quality Monitoring System with Power Consumption Optimization. – Cybernetics and Information Technologies, Vol. **20**, 2020, No 2, pp. 122-140.
25. S c h ü r h o l z, D., S. K u b l e r, A. Z a s l a v s k y. Artificial Intelligence-Enabled Context-Aware Air Quality Prediction for Smart Cities. – Journal of Cleaner Production, Vol. **271**, 2020, 121941.
26. H u, Z., et al. Real-Time Fine-Grained Air Quality Sensing Networks in Smart City: Design, Implementation, and Optimization. – IEEE Internet of Things Journal, Vol. **6**, 2019, No 5, pp. 7526-7542.
27. A t m a k u r i, K. C., Y. V. R. R a o. An IOT Based Novel Approach to Predict Air Quality Index (AQI) Using Optimized Bayesian Networks. – Journal of Mechanics of Continua and Mathematical Sciences, 2019.
28. P a n d y a, S., et al. Pollution Weather Prediction System: Smart Outdoor Pollution Monitoring and Prediction for Healthy Breathing and Living. – Sensors, Vol. **20**, 2020, No 18, p. 5448.
29. A m e e r, S., et al. Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities. – IEEE Access, Vol. **7**, 2019, pp.128325-128338.

30. M o u r s i, A. S., N. E l-F i s h a w y, S. D j a h e l, M. A. S h o u m a n. An IoT Enabled System for Enhanced Air Quality Monitoring and Prediction on the Edge. – Complex & Intelligent Systems, Vol. **7**, 2021, No 6, pp. 2923-2947.
31. S a i n i, J., M. D u t t a, G. M a r q u e s. ADFIST: Adaptive Dynamic Fuzzy Inference System Tree Driven by Optimized Knowledge Base for Indoor Air Quality Assessment. – Sensors, Vol. **22**, 2022, No 3, p.1008.

Received: 13.12.2022; Second Version: 24.05.2023; Third Verion: 01.08.2023; Accepted: 16.10.2023