

Industrial Centric Node Localization and Pollution Prediction Using Hybrid Swarm Techniques

R. Saravana Ram^{1,*}, M. Vinoth Kumar², N. Krishnamoorthy³, A. Baseera⁴, D. Mansoor Hussain⁵ and N. Susila⁶

¹Department of Electronics and Communication Engineering, Anna University, University College of Engineering Dindigul, Dindigul, 624622, India

²Department of Computer Science and Engineering, Anna University, University College of Engineering Dindigul, Dindigul, 624622, India

³Department of Computer Science and Engineering, Kongu Engineering College, Perundurai, 638060, India

⁴School of Computing Science and Engineering, VIT Bhopal University, Bhopal, 466114, India

⁵Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, 641008, India

⁶Department of Information Technology, Sri Krishna College of Engineering and Technology, Coimbatore, 641008, India

*Corresponding Author: R Saravana Ram. Email: rsaravanaram21@yahoo.com

Received: 10 July 2021; Accepted: 30 August 2021

Abstract: Major fields such as military applications, medical fields, weather forecasting, and environmental applications use wireless sensor networks for major computing processes. Sensors play a vital role in emerging technologies of the 20th century. Localization of sensors in needed locations is a very serious problem. The environment is home to every living being in the world. The growth of industries after the industrial revolution increased pollution across the environment. Owing to recent uncontrolled growth and development, sensors to measure pollution levels across industries and surroundings are needed. An interesting and challenging task is choosing the place to fit the sensors. Many meta-heuristic techniques have been introduced in node localization. Swarm intelligent algorithms have proven their efficiency in many studies on localization problems. In this article, we introduce an industrial-centric approach to solve the problem of node localization in the sensor network. First, our work aims at selecting industrial areas in the sensed location. We use random forest regression methodology to select the polluted area. Then, the elephant herding algorithm is used in sensor node localization. These two algorithms are combined to produce the best standard result in localizing the sensor nodes. To check the proposed performance, experiments are conducted with data from the KDD Cup 2018, which contain the name of 35 stations with concentrations of air pollutants such as PM, SO₂, CO, NO₂, and O₃. These data are normalized and tested with algorithms. The results are comparatively analyzed with other swarm intelligence algorithms such as the elephant herding algorithm, particle swarm optimization, and machine learning algorithms such as decision tree regression and multi-layer perceptron. Results can indicate our proposed algorithm can suggest more meaningful locations for localizing the sensors in the topology. Our proposed method achieves a lower root mean square value with 0.06 to 0.08 for localizing with Stations 1 to 5.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Keywords: Wireless sensor networks; node localization; industrial-centric approach; random forest regression; elephant herding optimization; swarm intelligence; pollution

1 Introduction

The genuine factors that act as major challenges in our environment are air quality issues, water pollution issues, and radiation pollution problems. A healthy society is the main sustainable development goal among countries worldwide. At present, advanced environmental monitoring systems are highly focused on developing technologies such as the internet of things (IoT), parallel computing, sensors, and distributed computing. Under these circumstances, this research article aims to achieve smart placement of sensors and investigate predicting smart environment pollution, such as factors involving monitoring of air quality level, water quality level, radiation pollution levels, and advanced monitoring over agricultural systems

Wireless sensor nodes are placed and scattered at all locations in the environment. The sensors are placed for various applications. Localizing or predicting where to place sensor nodes is important to focus on the sensing area accurately and reduce redundancy of sensing the same environment using parallel computing. In our research problem, we introduce industrial-centric node localization using a decision tree with a swarm intelligence algorithm. Manufacturing industries emit more pollutants. Therefore, placing more nodes in such areas with help of IoT is necessary, as is decreasing the nodes when moving away from industries. Owing to this architecture, the cost of sensors will be reduced and data prediction accuracy increases.

As the technological and information field grows, wireless sensor networks (WSN) collect huge data in and around the environment for various applications like military projects, metrological applications, medical fields, and security surveillance. Broadly enabled applications significantly introduce advancement in WSN. At present, the IoT concept uses WSN for data collection and processing in real time [1]. The collection sensor nodes with other inexpensive devices form the WSN infrastructure. This network is used to monitor and detects the environmental data for computing [2,3]. Collected data are sent to sink nodes, which are a destination for processing and storing the collected data and sent to the network for various usage and user applications. The WSN has many advantages such as parallel computing, deployment, communications, transferring data, and organizing data. Day by day, various challenges are experienced across WSNs. Further implementing the WSN has enormous challenges such as localizing nodes, area coverage, energy consumption, and time of transfer in sensors. Among these various challenges, node localization is a very important and basic problem to address.

Validation of sensor node localization is very important to collect data effectively. If we do not localize the sensors at correct or needed locations, we will miss information or misread data while processing without correct information. The nodes cannot be reached due to some climatic factors and unreachable environments, creating the problem of communication and network structure in deploying nodes in visible and needed places for future applications. The GPS system deploys sensor nodes in most places without any constraints, leading to more energy consumption by transferring data with high cost. Node localization can be defined as placing the sensor nodes in exact places and needed places. The triangulation technique is suggested in previous studies for node localization [4]; some methods used so far include arrival time [5], using signals of radio for localization [6–8], and arrival angles for choosing location [9].

The location of a GPS system can be precisely identified using the help of anchor nodes. The sensor nodes of GPS and non-GPS have their own localization methods based on range-based approaches [10]. These range-based algorithms identify the node distances with the help of angle measurement with

unknown nodes. Here, the position of the unknown nodes is only known. Furthermore, this problem was addressed using a triangulation technique to identify the coordinates. Topological data use range-free algorithms for localization to identify target nodes. Past research proved that these algorithms are not economically affordable to all users [11–13]. Sometimes, when focusing localization on indoor-based environments that have walls, humidity and temperature are totally different. Range-based localization cannot be used here. In this case, localization of nodes must be wirelessly done in areas within buildings.

The main goal in node localization is to place sensor nodes accurately at needed locations. In this article, WSN places sensors for pollution prediction. Previous research mainly focuses on cost-effective sensor node localization. Industries are commonly acknowledged to be the main sources of pollution in every location. Our contributions in this research paper are as follows.

1. Industrial areas are identified by pollution level using random forest regression techniques.
2. Once the location is identified, we use the elephant herding algorithm to decide the number of sensors and location in that industrial area.
3. Fringes of or places far away from industries do not need the number of sensors that we place near industrial areas.
4. This focus can reduce the cost of sensors by reducing extra or wasted sensor locations.

Node localization is also a NP-hard scenario to identify accurate location for the sensor nodes. In this article, we only attempted using industrial centric concepts.

This paper has been organized with five chapters. Chapter 2 provides a literature survey. Chapter 3 gives details of the proposed methodology and its implementation strategy. Chapter 4 evaluates the result by applying datasets. Chapter 5 concludes.

2 Literature Survey

The atmosphere is composed of different pollutant factors from industries, vehicles, and automobiles, such as PM_{2.5}, PM₁₀, ozone, sulfur dioxide, and nitrogen oxides. These pollutants are highly spread from manufacturing. However, pollution causes more damage to the environment and health that is irreparable. Many studies have proved the health effects occurring due to pollution, necessitating technological solutions to control and monitor the pollution concentration of various gases in the industries as well as in the surroundings. Various sensor solutions have been used for environmental quality assessment. At present, more sensor technologies are used, but they are expensive. The unavailability of low-cost techniques and quality of data makes processing ineffective data very difficult.

Next, different swarm intelligence algorithms used in node localization problems are reviewed. Solving NP-hard problems is very important. This case is also reviewed in the following section. Localizing the WSN is considered a real-life problem as an optimization task. Domains that focus on optimization problems are metaheuristic and artificial intelligence algorithms. Practically, NP-hard problems are addressed using metaheuristic solutions.

The nature-inspired metaheuristic algorithms can be classified into evolutionary and swarm intelligence. Genetic algorithm (GA) is a commonly used evolutionary algorithm for various NP-hard problems and WSN. The node localization uses GA with a range-free distributed algorithm for 3D WSN [14]. GA-based localization techniques provide accuracy with unknown sensor nodes in WSN [15]. The first swarm intelligence method described in our review work is particle swarm optimization (PSO) [16]. Flocks of fish and birds simulation are used in the search technique of PSO. This technique is highly applied in WSN node localization. Addressing the problems in WSN localization are addressed by swarm-based

topologies and variants of PSO for optimal solution [17]. Velocity PSO [18,19] improves the accuracy of the localization in the WSN. Hybrid PSO in [20] improves localization performance.

The artificial bee colony (BC) optimization algorithm is used in localization of NP-hard problems. In [21], BC localization is used in optimizing the nodes in the network. The firefly algorithm is an effective optimization algorithm in swarm intelligence techniques. It uses the light properties of the fireflies in addressing the localization. Many hybrid versions of the firefly algorithm are used in WSN localization [22]. The virtual node projection in the anchor nodes for analyzing target nodes for localization uses firefly concepts [23]. A novel swarm intelligence approach, namely, the monarch butterfly optimization (MB) method, is widely used in solving NP-hard problems. It was first introduced in 2015 by Wang et al. [24]. The MB algorithm is highly effective in optimizing NP solutions. WSN localization that uses MB produced accurate results. MB also addressed the multi-localization stage of WSN nodes in the network [25]. Moth search optimization is another swarm intelligent technique for WSN node localization strategy. It is highly appreciated using axis of photo and levy. It also proved its performance in solving benchmark problems [26]. Moth search is used in real-time problems like machine movement, drone control, and node localization in WSN [27]. The hybridized technologies in WSN have also been tested and implemented [28]. The 5G cloud computing environment [29] and adversarial networks generation [30] address next-generation demands for parallel computing. Low energy consumption [31] with image color recognition [32] are reconfigured for the network connectivity actions.

3 Proposed Hybrid RFR–EHO Methodology

Our goal is to place the nodes similar to IoT devices based on an industrial-centric environment. Placing more nodes in empty areas without any industries is a waste of sensor cost. In this proposed work, we use **random forest regression** to decide the number of nodes based on parallel processing of the decision from the algorithm indicating industrial or non-industrial areas. Then, the swarm intelligent technique called **elephant herding optimization** (EHO) is further used for node localization to place the number of sensors to increase the accuracy of data prediction. A previous study [33] proved that random forest regression predicts the air quality accurately with a low error rate and proved that among the various swarm intelligence algorithms in WSN for localization, EHO is efficient and robust to address WSN localization problems.

Our proposed work is smart environment pollution prediction on industries based on random forest regression for decision-making on the type of area and elephant herding to localize the sensors in that area. The overall workflow of the proposed work is diagrammatically presented in Fig. 1.

Fig. 1 illustrates the workflow of the proposed work to predict air pollution in industrial and non-industrial areas. The proposed work is divided into 4 phases, as follows:

1. Data gathering–air pollutants such as CO₂, SO₂, and PM.
2. Prediction of the area–Random forest regression–Industrial or non-industrial areas
3. Localization–EHO–Placing sensors on needed areas
4. Data management–Sensor result storage and alerts about the current environment status.

3.1 Data Gathering

The data that are used for this prediction model are gathered from the various heterogeneous nodes similar to IoT devices that are connected to the areas. Pollutants of industries are ozone gas (O₃), nitrogen dioxides (NO₂), carbon monoxides (CO), sulfur dioxides (SO₂), and particulate materials (PM). These pollutant data are collected and pre-processed because each pollutant format may vary. Pre-processing is the step to filter out unnecessary information and make the data ready for the next level of processing. In this preprocessing stage, the missing values are handled and the data are normalized.

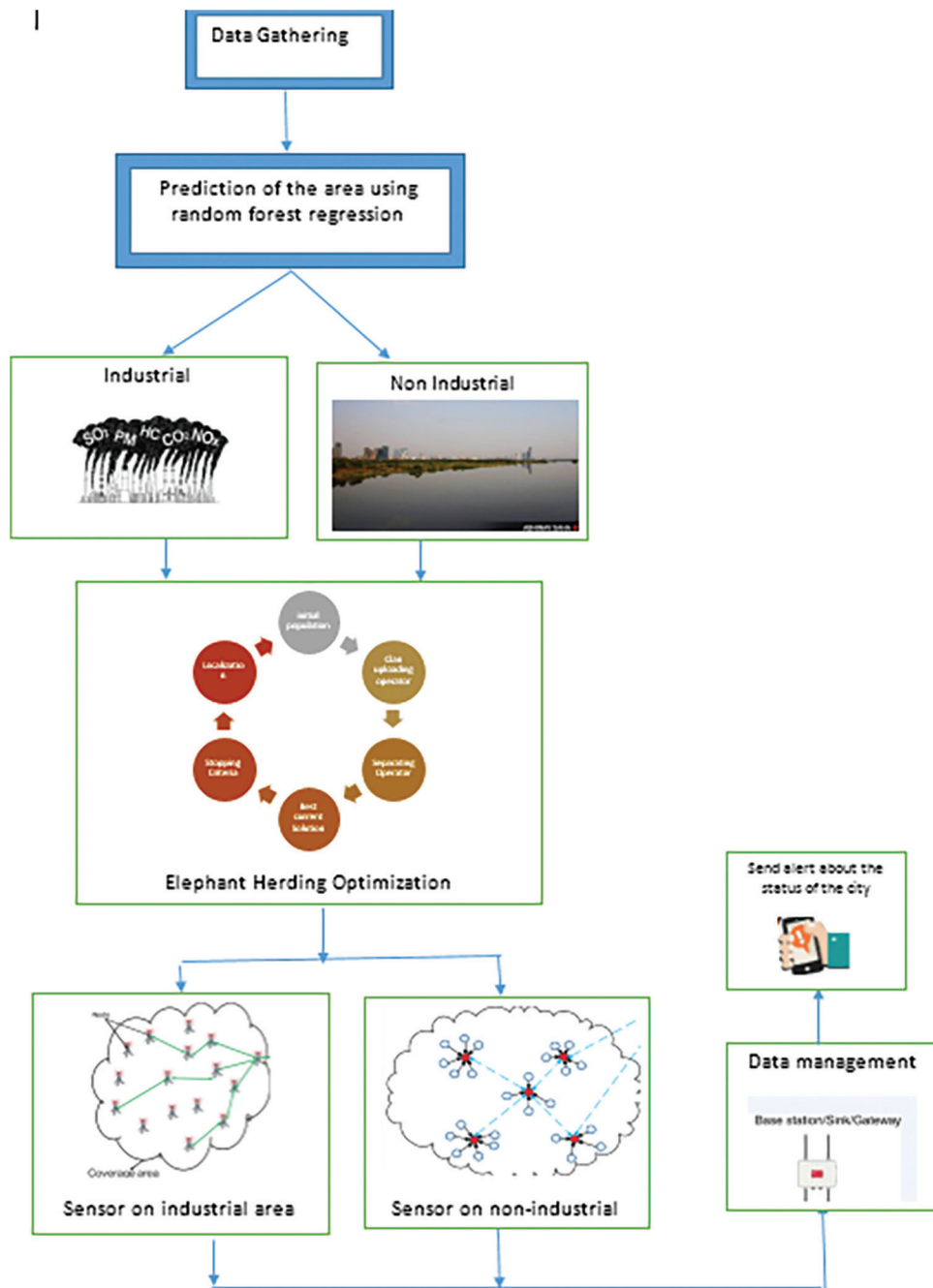


Figure 1: Proposed RFR-EHO working overview

Algorithm 1: Normalization of data

Input: Air pollutant data

Output: Normalized data in the range of [0,1]

Step 1: for each row

Step 2: //Split the data based on the collected area

(Continued)

 Algorithm 1: (Continued)

```

    Create file called area, and add concentration that collected in that area
Step 3:    // Handle missing values
           For each column
             If value of column=Not a Null
               Extract that row which contain that column
             End if
           End for
           End for
Step 4:    //Normalize the data
           For i = 1 to n // for each row
             For j = 1 to n //for each column
               Call min-max scaler (feature_range=(0, 1),*, copy=true) // from python sklearn
             End for
           End for
End for
Step 5: End
  
```

3.2 Prediction Analysis (Random Forest Regression)

Random forest is a decision tree algorithm that uses a tree structure. Each tree is generated from the data sample from the training data. When generating the tree, the random subset of the selected features is considered the best split among all the other alternatives. The process of random forest regression is shown in Fig. 2. Owing to random choice, the bias may increase and variance may decrease. In this proposed work, on the basis of this random forest regression, the polluted area that has a high air quality index (AQI) is considered an industrial area and that with a low AQI value is considered a non-industrial area.

The AQI value is calculated using Eq. (1) from China's Environmental Protection Ministry. AQI is the maximum value of the IAQI of one air pollutant p.

$$AQI_p = \frac{IAQI_{H_i} - IAQI_{L_o}}{BP_{H_i} - BP_{L_o}} (C_p - BP_{L_o}) + IAQI_{L_o} \quad (1)$$

where

C_p – mass concentration of each air pollutant p

BP_{H_i} – high value of pollutant's concentration limit

BP_{L_o} – low value of pollutant's concentration limit

$IAQI_{H_i}, IAQI_{L_o}$ – corresponding value of BP_{H_i} and BP_{L_o}

The corresponding AQI with air pollution level is categorized in Tab. 1.

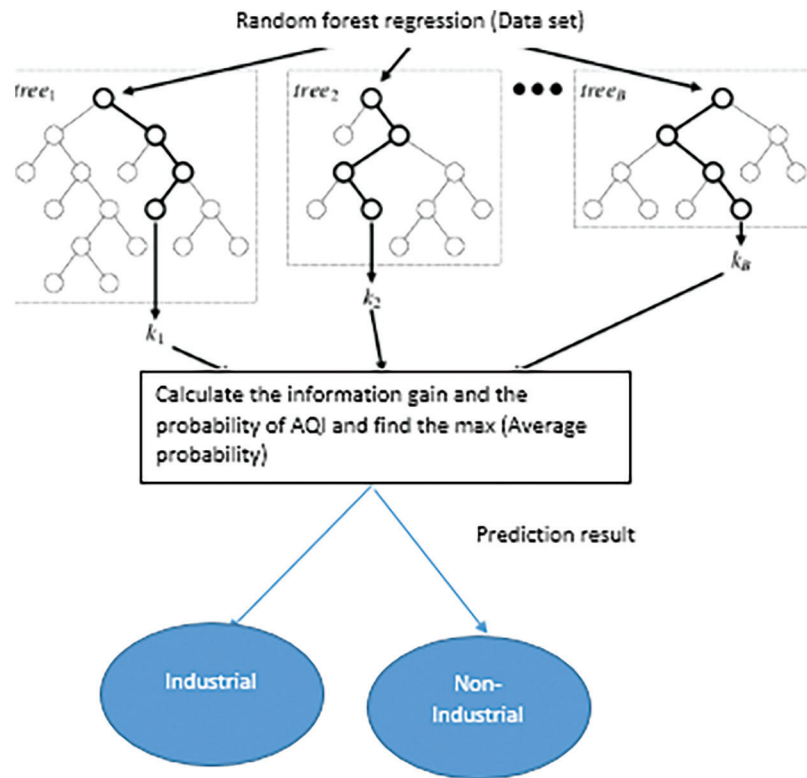


Figure 2: Random forest regression prediction of the proposed work

Table 1: AQI level for area prediction

AQI	Air pollution area category
0–50	Not polluted
51–150	Low pollution
151–200	Moderate pollution
201–300+	Severe pollution

Algorithm 2: AQI prediction using Random forest regression with EHO

Input: normalized data set with features labeled with AQI level, unlabeled dataset UD, trees quantity T, features quantity m

Output: prediction of the area using AQI level

Step 1: For T trees

Step 2: Random selection of m features from s

Step 3: For feature m in each tree node

To split the data set, Calculate information gain as

$$Entropy(c) = - \sum_{i=1}^k p(c_i) \log_2 p(c_i)$$

(Continued)

 Algorithm 2: (Continued)

$$p(c_i) = \frac{N_i}{\sum_{i=1}^k N_i}$$

$$\text{Information Gain (IG } (f_i)) = Entropy(c) - \sum_{j=1}^n \frac{|f_i^j|}{|f_i|} \times Entropy(f_i^j) \quad (3)$$

where

c_i – AQI level of i

k – total number of AQI levels

N_i – quantity of records

f_i – records of i^{th} level of T

f_i^j – records of j^{th} node of i^{th} level of T

n – number of nodes in the level

Step 4: Split the data set node using $\max(\text{IG})$

Step 5: Remove selected features from S

Step 6: Input UD into T . The final probability of the i^{th} level AQI is calculated as

$$p'(c_i) = \frac{1}{T} \sum_{k=1}^T p(c_i) \quad (4)$$

Step 7: Predict the area based on AQI level using

$$PC'(i) = \max(p'(c_i)) \quad (5)$$

On the basis of the air pollution quality index, the polluted area is categorized as not polluted, low, medium, and high. An AQI level more than 200 is considered the industrial area. For our proposed work area, industry classification is greatly needed for sensor placement for better prediction because greatly polluted areas need more sensors for accurate prediction.

3.3 EHO Algorithm for Node Localization

The localization procedure in WSN consists of anchor nodes and unknown nodes. The process has two phases. First is the ranging phase where the algorithm finds the distance between anchor nodes and unknown nodes. In the second phase, a sensor is fitted based on the position of the nodes based on angle of arrival, round trip time, radio signal strength, time of arrival, and time difference of arrival. The principle of localization to place wireless sensor consists of M sensors located in N unknown nodes using the information of the location of $M-N$ anchors and transmission range. If a sensor is placed within the transmission range of three or more anchors, then it is considered localized. In this paper, the EHO algorithm is used for localization problems. This algorithm has been proven to be the best method for localization in WSN.

EHO has been used to solve global optimization problems [34,35]. They proposed a heuristic search that is based on the co-existence of the elephants in the clan that are guided by the leader or the matriarch, which is the oldest female in the clan. The other members of that clan are females and calves, while the male elephants leave the clan and live separately. The male elephants can also communicate with the clan. This concept can be used in node localization in WSN. On the basis of structural differences, this EHO can have two environments such as elephants living under the guidance of the matriarch and males living separately but still in communication with the clan. These environments are used as a separating operator.

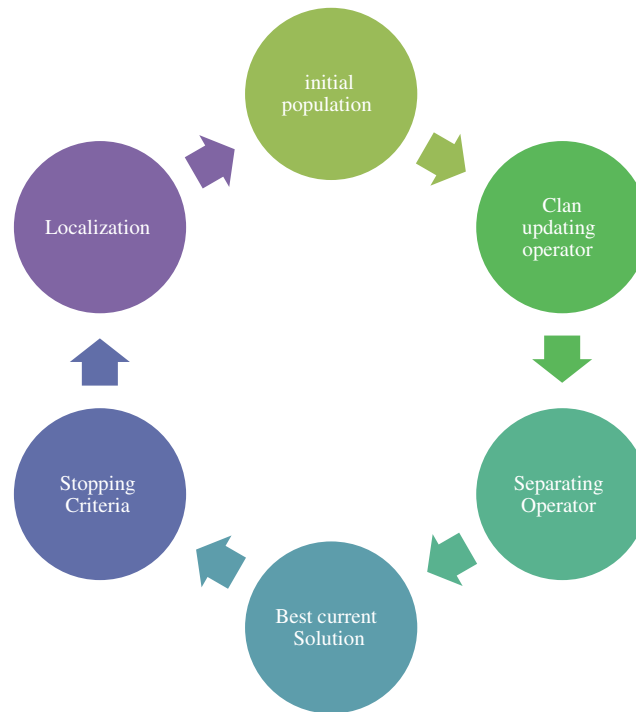


Figure 3: EHO localization structure

The general structure of the EHO algorithm is shown in Fig. 3. In EHO, each possible solution of clan c_i is updated with the current position and matriarch c_i using the updating operation. The population difference is updated using a separating operator. Initially, the population is divided into n clans.

Algorithm 3: EHO localization

Input: Predicted features of the pollutant area using algorithm 2

Output: Best possible solution of the positions to place sensors

Step 1: Initialization: generate the predicted features from algorithm 2 as population; divide the population into ‘ n ’ clans; calculate fitness of each individual of the population; initialize the generation counter as $c = 1$ and maximum gen (MG)

Step 2: while $c < \text{MG}$ do

Step 3: Arrangement of the solutions based on fitness value

Step 4: For all clans c_i do

Step 5: For all solution j in the clan c_i do

Step 6: Compute updating operation of the clan using Eq. (6)

$$x_{newc(i,j)} = x_{c(i,j)} + \alpha \times (x_{bestci} - x_{c(i,j)}) \times r \quad (6)$$

Step 7: Select better solution between $x_{c(i,j)}$ and $x_{new,c(i,j)}$

Step 8: Update the fittest value of $x_{best,ci}$ and generate new population $x_{new,c(i,j)}$

Using Eq. (7)

(Continued)

 Algorithm 3: (Continued)

$$x_{newc(ij)} = \beta \times x_{centercid} \quad (7)$$

$$x_{centercid} = \frac{1}{n_{ci}} \times \sum_{j=1}^d x_{cijd} \quad (8)$$

 Step 9: Select better solution between $x_{best,ci}$ and $x_{new,c(i,j)}$

End for

End for

 Step 10: For all clan ci in the population do

Replace the worst solution using the separating operator as Eq. (9)

$$x_{worstci} = x_{min} + (x_{max} - x_{min} + 1) \times rand \quad (9)$$

End for

Step 11: Compute the population and fitness value

Step 12: End while

Step 13: Return the best possible solution

 where, $x_{c(i,j)}$ – old position of j in clan ci
 $x_{new,c(i,j)}$ – new position of j in the clan ci
 $x_{best,ci}$ – best solution of clan ci
 α, β, r and $rand \in [0, 1]$ – scale indicator, factor on $x_{center,ci}$ and random variable

 x_{max} and x_{min} – upper and lower bound position of the individual

 $x_{worst,ci}$ – individual with the worst fitness value in the clan ci

On the basis of this EHO, the multi-stage WSN localization for our proposed work is shown in Fig. 4. In multi-stage localization, unknown sensor nodes with three or more anchor neighbors can be localized. In single-stage localization, the unknown node that has more neighbors can be localized. The multi-stage localization of the industrial area results in more sensors placed on highly polluted positions for accurate prediction of air pollution levels.

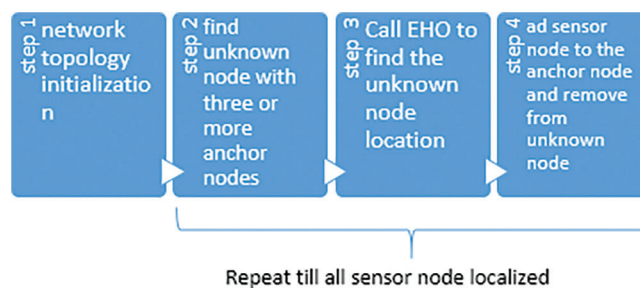


Figure 4: Multi-stage localization of sensor nodes using EHO

3.4 Data Management

Storing and managing the localized data is an important step for analysis in real-time processing. Given that our proposed work is for smart environment pollution prediction, the data must be stored and processed rapidly. Our processed data are stored in an HDFS system. The stored information is then communicated to

the application with an interface. The communication is frequently sent to the end user about the levels of polluted areas. This communication will announce all the information about the polluted areas so that people can monitor pollution there and make precautions and decisions accordingly.

Our proposed work for finding the air polluted area based on the air pollutant index with random forest regression tree will categorize the area affected by pollution as industrial and non-industrial areas. This process will help to localize the sensor count as needed according to air pollution using EHO localization. Such localization will improve the efficiency and prediction accuracy because more sensors will be placed in the most affected areas, leading to improved prediction with quick response time. Thus, authorities will take necessary actions to prevent that. The aim of this work is to place more sensors in needed areas than in normal areas, creating a smart environment to predict air pollution with fast response time and high accuracy.

4 Results and Discussion

4.1 Data Description

The experimental analysis of our proposed smart air pollution prediction environment has been tested using data from the KDD Cup 2018, which contains the names of 35 stations with concentrations of air pollutants such as PM, SO₂, CO, NO₂, and O₃. The raw dataset has missing values and is not normalized. During the data-gathering phase of our proposed work, the raw data are preprocessed to obtain normalized data [36]. The normalized data set is shown in Tab. 2. The algorithms are implanted using Python programming. The preprocessing stage is done by Panda. The evaluation metrics are calculated using sklearn in Python.

Table 2: Data set description after pre-processing

	PM	SO ₂	CO	NO ₂	O ₃
Count	200.00	200.00	200.00	200.00	200.00
Mean	134.376	27.180	17.205	15.788	11.212
Standard Deviation	123.790	56.373	20.671	11.056	2.788
Min	4.600	0.200	0.200	2.000	2.000
Max	561.000	208.000	79.000	61.000	37.000

4.2 Experimental Evaluation

To evaluate the proposed work with the simulation data set, the error between actual and predicted values are calculated using different types of error measure such as mean absolute error (MAE), root mean square error (RMSE), and accuracy. MAE is the measure of the average magnitude of the errors in the set of predictions. MAE is the measure of the difference between the actual and predicted values using Eq. (7). RMSE is the measurement of the average of the squared differences between the actual and predicted value using Eq. (8).

$$MAE = \frac{1}{n} \sum_{j=1}^n (y_j - y'_j) \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - y'_j)^2} \quad (8)$$

n-number of observations, y_j - actual value, y'_j - predicted value

The data set is split by station and saved as a separate file to identify the stations with high air pollution. Each station consists of these six types of pollutant concentrations. Thus, the data set is split into 35 stations with 6 features and 8,886 records. For our evaluation, we consider five stations core stations and use our proposed algorithm to predict air pollution in a smart way. The error value for the proposed method in these five stations are calculated as shown in [Tab. 3](#), with the illustrated chart in [Fig. 5](#). The MAE, RMSE, and accuracy of the proposed smart prediction method for the five stations are computed and stated with minimum MAE and minimum error value. To analyze the performance of the proposed scheme, it is compared with existing air prediction techniques such as decision tree regression (DTR) [33], hybridized elephant herding optimization (HEHO) [5], multi-layer perceptron (MLP) [33], and particle swarm optimization (PSO) [36].

Table 3: Error value of proposed scheme

Stations	MAE (%)	RMSE
ST1	11.2	0.0621
ST2	10.5	0.07352
ST3	14.83	0.0523
ST4	12.22	0.03828
ST5	10.83	0.0823

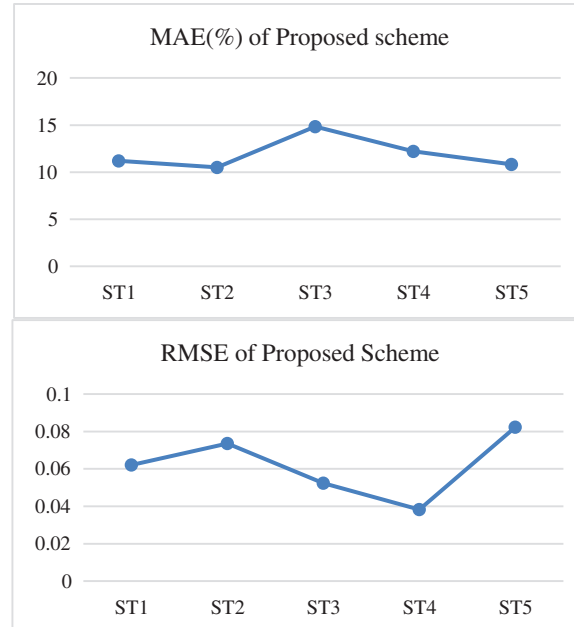


Figure 5: Proposed scheme error value calculation

The performance evaluation of the proposed algorithm with the existing algorithms is shown in [Tab. 4](#). The observed result of the table values shows that our proposed algorithm obtains a low error value compared with other existing algorithms. The proposed algorithm obtains an average of 11.916% of MAE and 0.0617 of RMSE, representing a high level of prediction accuracy. The graphical representation of these

data values is shown in Figs. 6 and 7. The proposed method has lower representation than do other methods. The next best method to the proposed method is the HEHO algorithm with an average of 16.27% of MAE, which is close to MLP, and 0.1233 of RMSE.

Table 4: Performance evaluation of proposed work

Stations	DTR		HEHO		MLP		PSO		Proposed hybrid RFR-EHO	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
ST1	17.32	0.142	10.23	0.0972	14.93	0.21	15.34	0.24	11.2	0.0621
ST2	15.34	0.163	17.56	0.173	12.38	0.293	18.35	0.192	10.5	0.07352
ST3	16.92	0.134	19.34	0.0893	17.35	0.234	18.25	0.153	14.83	0.0523
ST4	14.82	0.103	16.83	0.124	15.39	0.123	19.38	0.0937	12.22	0.03828
ST5	12.542	0.192	17.39	0.133	13.59	0.162	16.34	0.154	10.83	0.0823

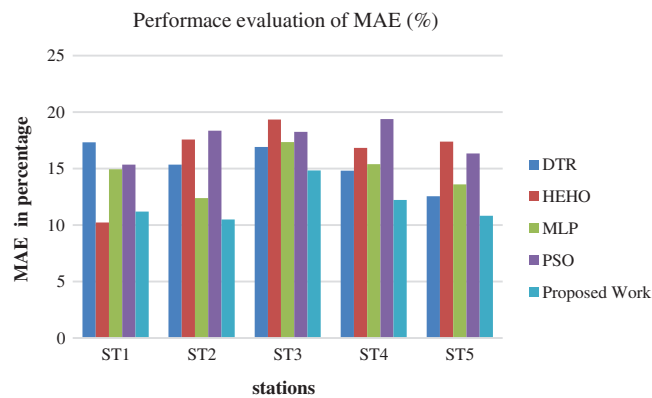


Figure 6: Performance evaluation of MAE of the proposed algorithm with existing methods

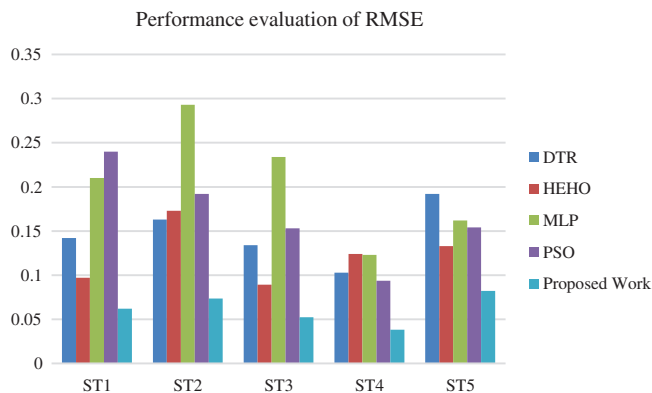


Figure 7: Performance evaluation of RMSE of the proposed algorithm with existing methods

The proposed algorithm performs better in prediction and localization. This method results in a smart environment with a smart air pollution prediction system with quick response. The method also reduces the cost of the sensors that are unnecessarily placed in less polluted areas.

5 Conclusion

The present work focuses on localizing nodes and reducing sensor cost by accurate placement using the using swarm intelligence algorithm. To achieve efficiency, we first implement the random forest regression technique to classify industrial areas. Then, the classified areas are localized using elephant herding techniques. The work has two advantages in deciding the number of sensors to be placed and solving node localization. The accuracy of sensor node localization is achieved by working with target nodes. WSN is very important in day-to-day life. Sensor information is important to various information processing applications. Localizing the nodes and knowing how many sensors are to be placed in what type of location is very important for economical computation. This process also helps to avoid redundancy in sensing the same locations in the network. Our approach overcomes all problems by using a hybridized solution for various problems. The performance and error rates are computed for the proposed algorithm, with a mean error value of less than 1. Existing techniques such as PSO, EHO, DTR, and MLP are compared with the proposed RFR–EHO approach. The performance of localization is markedly higher than all existing techniques. In our article, we focus only on industrial areas, but urban areas also have high pollution. The future scope of this article is to use an industrial- and urban-centric approach using deep learning with the swarm intelligence technique. Highly proven algorithms will be used in this approach.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] D. E. Boubiche, A. K. Pathan, J. L. loret, H.Zhou, S.Hong *et al.*, “Advanced industrial wireless sensor networks and intelligent IoT,” *IEEE Communications Magazine*, vol. 56, no. 2, pp. 14–15, 2018.
- [2] I. F. Akyildiz and M. C. Vuran, “Wireless sensor networks”, in *Series Communications and Networking*, 1st ed., Springer, 2010.
- [3] P. Rawat, K. D. Singh, H. Chaouchi and J. M. Bonnin, “Wireless sensor networks: A survey on recent developments and potential synergies,” *The Journal of Supercomputing*, vol. 68, no. 1, pp. 1–48, 2014.
- [4] S. Tomic and M. Beko, “Target localization via integrated and segregated ranging based on RSS and TOA measurements,” *Sensors*, vol. 19, no. 2, pp. 230–250, 2019.
- [5] I. Strumberger, M. Beko, M. Tuba, M. Minovic and N. Bacanin, “Elephant herding optimization algorithm for wireless sensor network localization problem,” in *Proc. DoCEIS*, Costa de Caparica, Portugal, pp. 175–184, 2018.
- [6] M. Garcia, J. Tomas, F. Boronat and J. Lloret, “The development of two systems for indoor wireless sensors self-location,” *Ad Hoc Sensor. Wireless. Networks*, vol. 8, no. 3, pp. 439–459, 2009.
- [7] S. Tomic, M. Beko, R. Dinis, M. Tuba and N. Bacanin, “A CRB derivation for RSS-AoA localization,” *RSS-AoA-based Target Localization and Tracking in Wireless Sensor Networks*, vol. 19, no. 2, pp. 87–92, 2017.
- [8] S. Tomic, M. Beko, R. Dinis, M. Tuba and N. Bacanin, “Bayesian methodology for target tracking using combined RSS and AoA measurements,” *Physical Communication*, vol. 25, pp. 158–166, 2017.
- [9] S. Tomic, M. Beko, R. Dinis and J. P. Gomes, “Target tracking with sensor navigation using coupled RSS and AOA measurements,” *Sensors*, vol. 17, no. 11, pp. 2690–2712, 2017.
- [10] S. Abdollahzadeh and N. J. Navimipour, “Deployment strategies in the wireless sensor network: A comprehensive review,” *Computer Communications*, vol. 91, pp. 1–16, 2016.

- [11] J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley *et al.*, “A theory of network localization,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 12, pp. 1663–1678, 2006.
- [12] S. Maheswaran, S. Sathesh, G. Saran and B. Vivek, “Automated coconut tree climber,” in *Proc.I2C2*, Coimbatore, India, vol. 4, pp. 184–189, 2017.
- [13] B. Vivek, S. Maheswaran, P. Keerthana, S. Sathesh, S. Bringeraj *et al.*, “Low cost raspberry Pi oscilloscope,” in *Proc.I2C2SW*, Erode, India, pp. 386–390, 2018.
- [14] Q. Han, Z. Xiong and K. Zhang, “Research on trajectory data releasing method via differential privacy based on spatial partition,” *Security and Communication Networks*, vol. 2018, pp. 1–14, 2018.
- [15] G. Marques and R. Pitarma, “An indoor monitoring system for ambient assisted living based on internet of things architecture,” *International Journal of Environmental Research and Public Health*, vol. 13, no. 11, pp. 1152–1166, 2016.
- [16] F. Salamone, L. Belussi, L. Danza, T. Galanos, M. Ghellere *et al.*, “Design and development of a wearable wireless system to control indoor air quality and indoor lighting quality,” *Sensors*, vol. 17, no. 5, pp. 1021–1038, 2017.
- [17] P. Srivatsa and A. Pandhare, “Indoor air quality: IoT solution,” in *Proc. NCPCL*, Coimbatore, India, 2016.
- [18] V. Wittstock, L. Scholz, B. Bierer, A. O. Perez, J. Wöllenstein *et al.*, “Design of a LED-based sensor for monitoring the lower explosion limit of methane,” *Sensors and Actuators B: Chemical*, vol. 247, pp. 930–939, 2017.
- [19] C. Wong, M. MatJafri, K. Abdullah, H. Lim and K. Low, “Development of air quality monitoring remote sensor using a digital SLR camera,” in *Proc. Aerospace Conf.*, MT, USA, pp. 1–6, 2008.
- [20] S. G. Narasimhan and S. K. Nayar, “Contrast restoration of weather degraded images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 713–724, 2003.
- [21] G. Papadopoulos and G. Mair, “Amplitude and phase study of the photoacoustic effect,” *Journal of Physics D: Applied Physics*, vol. 25, no. 4, pp. 722–801, 1992.
- [22] J. Kneer, A. Eberhardt, P. Walden, A. Ortiz Pérez, J. Wöllenstein *et al.*, “Apparatus to characterize gas sensor response under real-world conditions in the lab,” *Review of Scientific Instruments*, vol. 85, no. 5, pp. 055006–055018, 2014.
- [23] R. Li and R. Kang, “Research on failure rate forecasting method based on ARMA model,” *Systems Engineering and Electronics*, vol. 30, no. 8, pp. 1588–1591, 2008.
- [24] C. M. Rocco, “Singular spectrum analysis and forecasting of failure time series,” *Reliability Engineering and System Safety*, vol. 114, pp. 126–136, 2013.
- [25] M. das Chagas Moura, E. Zio, I. D. Lins and E. Drogue, “Failure and reliability prediction by support vector machines regression of time series data,” *Reliability Engineering & System Safety*, vol. 96, no. 11, pp. 1527–1534, 2011.
- [26] K. Xu, M. Xie, L. C. Tang and S. L. Ho, “Application of neural networks in forecasting engine systems reliability,” *Applied Soft Computing*, vol. 2, no. 4, pp. 255–268, 2003.
- [27] X. Wang, J. Wu, C. Liu, S. Wang and W. Niu, “A hybrid model based on singular spectrum analysis and support vector machines regression for failure time series prediction,” *Quality and Reliability Engineering International*, vol. 32, no. 8, pp. 2717–2738, 2016.
- [28] R. S. Ram, M. V. Kumar, S. Ramamoorthy, B. S. Balaji and T. R. Kumar, “An efficient hybrid computing environment to develop a confidential and authenticated IoT service model,” *Wireless Personal Communications*, vol. 117, no. 4, pp. 2903–2927, 2020.
- [29] M. V. Kumar, R. S. Ram, B. Amutha and M. Kowsigan, “An improved migration scheme to achieve the optimal service time for the active jobs in 5G cloud environment,” *Computer Communications*, vol. 160, pp. 807–814, 2020.
- [30] S. R. Ram, B. Subramanian, N. Bacanin, M. Zivkovic and I. Strumberger, “Speech enhancement through improvised conditional generative adversarial networks,” *Microprocessors and Microsystems*, vol. 79, pp. 103281–103296, 2020.

- [31] R. S. Ram, A. G. Saminathan and S. A. Prakash, "An area efficient and low power consumption of run time digital system based on dynamic partial reconfiguration," *International Journal of Parallel Programming*, vol. 48, no. 3, pp. 431–446, 2020.
- [32] R. S. Ram, S. A. Prakash, M. Balaanand and C. Sivaparthipan, "Colour and orientation of pixel based video retrieval using IHBM similarity measure," *Multimedia Tools and Applications*, vol. 79, no. 15, pp. 10199–10214, 2020.
- [33] S. Ameer, M. Ali Shah, A. Khan, H. Song, C. Maple *et al.*, "Comparative analysis of machine learning techniques for predicting air quality in smart cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019.
- [34] Y. Liu, Z. Yang, X. Wang and L. Jian, "Location, localization, and localizability," *Journal of Computer Science and Technology*, vol. 25, no. 2, pp. 274–297, 2010.
- [35] G. G. Wang, S. Deb and L. d. S.Coelho, "Elephant herding optimization," in *Proc. ISCBI*, IEEE, Bali, Indonesia, pp. 1–5, 2015.
- [36] S. A. Janabi, M. Mohammad and A. A. Sultan, "A new method for prediction of air pollution based on intelligent computation," *Soft Computing*, vol. 24, no. 1, pp. 661–680, 2020.