Global NEST Journal, Vol 25, No 10, pp 47-55

A deep learning model and optimization algorithm to forecasting environment monitoring of the air pollution

Jayamala R.[1], Shanmugapriya N.[2], Lalitha K.[3] and Vijayarajan P.[4]

[1]Department of Computer Science and Engineering, University College of Engineering, BIT Campus, Anna University, Trichy, India

[2]Department of Computer Science and Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Samayapuram, Trichy, India

[3]Department of Information Technology, Panimalar Engineering College, Chennai, India

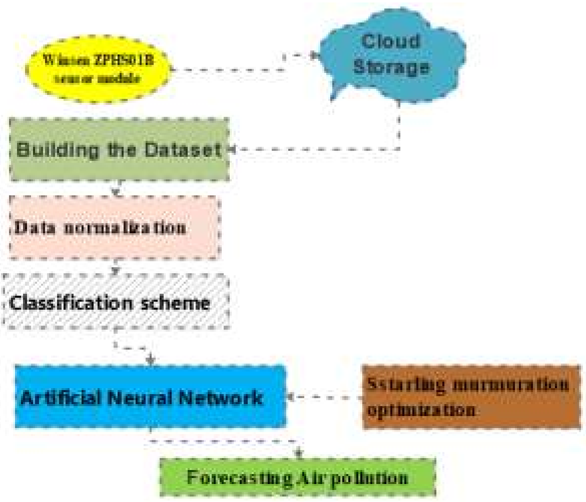
[4]Department of Electrical and Electronics Engineering University College of Engineering BIT Campus Tiruchirapalli, India

Received: 26/01/2023, Accepted: 15/09/2023, Available online: 24/09/2023

*to whom all correspondence should be addressed: e-mail: jayamalar546@gmail.com

https://doi.org/10.30955/gnj.004759

Graphical abstract



Abstract

Air pollution monitoring is becoming increasingly important, with an emphasis on the effects on human health. Because nitrogen dioxide (NO_2) and sulphur dioxide (SO_2) are the principal pollutants, many models for forecasting their potential harm have been created. Nonetheless, making precise predictions is nearly impossible. The prediction of air pollution enables researchers to understand how pollution affects human health. Deteriorating air quality can lead to respiratory diseases such as lung cancer and asthma. The effect of pollution on environmental degradation can also be predicted and reductions can be detected in the ozone layer. This study also focuses on and promotes the development of smart city environments by obtaining influential pollutants that affect the air, thereby reducing the source of specific pollutants An Artificial Neural Network (ANN) model is used as a forecast the pollution and the starling murmuration optimization (SMO) procedure is used to optimise the Artificial Neural Network strictures to achieve a lower forecasting error. Furthermore, in this research work, we used real time dataset as we have used Winsen ZPHS01B sensor module to collect the data, which is stored in cloud platform. After the composed data is used to train and test, after this process we will evaluate the results. To assess the performance of the suggested model. Furthermore, the perfect has been tested using two alternative kinds of input parameters: type as, which contains various lagged values of variables (NO_2 and SO_2), and type as, which only includes lagged values of the yield variables. The collected findings suggest that the projected model is more precise than existing joint forecasting benchmark models when different network input variables are considered.

Keywords: Artificial neural network, sulfur dioxide, deep learning model, air pollution forecasting

1. Introduction

Because of its impact on well-being, air quality is a major source of concern. Although decreasing air quality is a major problem in underdeveloped countries, it can also have substantial repercussions in affluent ones. Rendering, the level of indoor air pollutants may be 100 times higher than the level of outside air pollutants. The Index of Air Quality (IAQ) prediction system contributes to smart environments, where advanced sensor technologies can be employed to provide people with healthy living circumstances. There has been an increasing interest in using ANN models to forecast environmental air pollution in recent years. The deterioration of urban air quality is to blame for chronic diseases and the early mortality of susceptible groups and individuals (Zhang et al., 2022). As a result, there is an urgent need for policymakers and urban planners to propose quick and simple remedies to mitigate the detrimental effects of air pollution (Snežana et al., 2023). ANN has been effectively deployed in a variety of short-term and long-term prediction applications in recent years (Zho et al., 2019).

Furthermore, practitioners and researchers are increasingly inclined to use data-driven methods like artificial neural networks to replace traditional physics-based methods like the Urban Airshed Model (UAM) (Thiruppathi and Vinoth Kumar, 2011; Yan et al., 2021), Chemical Weather Research and Forecast model, and Community Multi-scale Air Quality Model (CMAQ). This observation can be attributed to the fact that deterministic methods are sensitive to multiple factors such as the scale and quality of the parameters involved, as well as being computationally intensive due to the fact that these methods rely on large databases with multiple input parameters (Chang et al., 2020). The use of ANN modelling, on the other hand, does not necessitate a thorough grasp of the dynamic link between air pollution attentiveness levels and other variables. In recent years, the public has had more access to powerful and less difficult computing tools for the development and deployment of ANN and their training algorithms (Mujeeb et al., 2019). Neural networks are utilised in stock exchange prediction, handwriting recognition, machine diagnostic, portfolio management, credit rating, speech recognition, intelligent searching, process modelling and control forecasting, and sales professional travel difficulties. The neural network is critical in prediction and categorization.

Several Environmental Science and Engineering researchers have worked on developing models to forecast the quantities of gases that cause air pollution. Artificial neural networks' high capabilities, such as their flexible structure and use of dynamic learning algorithms, encourage the adoption of these intelligent systems in this domain. To anticipate air pollution, most environmental researchers employ Artificial Neural Network (ANN) models. Numerous studies have been conducted on the nature and dynamic behaviour of pollutants, as well as their emission, spread, and impacts (Qi et al., 2019; Rao et al., 2019; Kavitha et al., 2023). Based on legislation and existing air quality regulations related to the meteorological variable, the ANN model was used to predict Particulate Matter pollution concentrations on a daily basis. The AQI is very useful in monitoring regular air quality and is thus used to raise public awareness in urban and rural areas. AQI can also be used to assess the influence of current air quality on public health (Cabaneros et al., 2019). Respirable Suspended Particulate Matter are a few air pollutants that are damaging to human health (RSPM). The high level of air pollutants has negative impacts on human health, including an increased risk of heart failure, asthma, giddiness, and other health difficulties. As a result, governments at the municipal and state levels are making concerted efforts to accurately assess and anticipate the air quality index, thereby reducing these threats to public health (Tao et al., 2019; Shamshirband et al., 2018).

2. Literature survey

Several Machine Learning and Artificial Neural Network techniques have been introduced to resolve issues on air pollution prediction and forecasting in the environment of smart cities. However, to the best of our knowledge, very few systematic literature reviews have been performed on the use of ANN in predicting air quality for smart cities.

Xiao Feng et al., (2015) combined model employing air mass route and wavelet transformation was used to recover the accuracy of ANN prediction for the daily average of the Particulate Matter (PM2.5) pollutant attentiveness. Data from 13 air pollution monitoring stations in Beijing, Tianjin, and Hebei Province of China were used for forecasting. The model utilized the Back Propagation Neural Network Multi- Layer Perceptron (BPNN-MLP) for daily weather forecasts using their respective pollutant forecast variables as input. The study successfully verified the proposed technique for almost 1 year (September 2013 to October 2014). The results indicated that the create the geographic model and wavelet transformation effectively improved the prediction accuracy of PM2.5, thus being recommended as a suitable model for air quality prediction.

Lu Bai et al. (2018) in this study, air pollution prediction methods were broadly classified into statistical forecasting approaches, numerical forecasting methods. Apart from these methods, hybrid methods were also applied to enhance air pollution forecasting. Additionally, this study investigated the theoretical aspects of prediction techniques and their implementation, together with their advantages and disadvantages. The study concluded that artificial intelligence methods such as the neural network method achieved enhanced forecasting and computing capabilities, even for non-linear and unbalanced datasets.

H. Haviluddin et al., (2018) developed a SOM model for spatial interpolation, whereby Learning Vector Quantization (LVQ) was used to optimize the SOM to obtain a better distance function between the sample datasets. n this study, datasets of land value zones in Indonesia, East Kalimantan, and Samarinda were utilized. SOM generated the intra-layer final weights to identify the centroid clusters and optimized them using LVQ. The results indicated that SOM-LVQ achieved a better distance variance value in comparison to traditional SOM. However, as the number of clusters increases, the mean absolute percentage error was found to be higher.

Rajeev Tiwari et al., (2019) employed the use of ANN to predict air quality using a constrained dataset derived from the United States on daily basis from 2008 to 2017 and processing data containing noise and errors. The predicted AQI was verified by the Auto-Regressive prediction model. The estimated value of Mean Square Error (MSE) validated the capabilities of the developed model, thus proving its effectiveness.

Hamed K, Qi L et al., (2019) evaluated three models for forecasting the PM2.5 pollutant. These models consisted of Multiple value of 8.91 μgm-3 and Mean Absolute Error (MAE) value of 6.21 μgm-3 was obtained using the LSTM model. Therefore, the LSTM model was superior to the other two models and was thus deemed effective for the prediction and control of air pollution. In addition, the

LSTM model achieved 80% variability (R2 = 0.8) at PM2.5 concentration and predicted a 75% pollution level. Furthermore, the results showed that the MART model estimated PM2.5 better than the DFNN model.

Z. Zhao, et al., (2020) developed a hybrid synchronous network model known as Collaborative Evolutionary Reinforcement Learning (CERL) to manage temporary serial data for the prediction of air characteristics over numerous hours. CERL aims to copy multiple basic models into the data. As compared with other basic models, the CERL model was shown to be superior. MAPE was used for 1, 3, 5, and 8 step predictions that were regarded as validation parameters. However, the proposed model was only based on hourly predictions and lacked high accuracy for long-term predictions.

Zhou, et al., (2021) employed a neural network and enhanced it with a non-linear neural network for time series forecasting. Two models were proposed, with one model based on previous AQI and the other based on meteorological statistics. These models were initially used to estimate the pollution attentiveness and subsequently used to calculate the Air Quality Index (AQI). The experimental results to verify the effectiveness of the model were evaluated based on band accuracy. The method achieved the accuracy of all records, which was over 81%. However, for the analysis of the estimated value of an individual pollutant, the study concluded that the PM10 estimation technique based on neural networks was not optimal.

3. Problem statements

Environmental air pollution shortens the lives of organisms including humans, animals, and plants. As a consequence, the natural air quality is affected, and its impact is increasingly affecting daily activities. Several factors such as industrialization, urbanization of farmland, and forest civilization directly affect air pollution. From a social perspective, the air quality must be analyzed to take necessary preventive and corrective measures to ensure a clean and sustainable living environment. Different mathematical models were identified and applied to understand the components of air pollution. Air pollution analysis involves two main activities, namely pollutant composition, and pollution prediction. To predict the level or condition of air pollution, it is essential to process and use historical data that measures the parameters by considering the availability of large amounts of datasets to distinguish the types and degrees of relationships for proper and effective extraction of information. Most of the spatial data collected, such as air quality, are scattered in nature, with different pollution levels observed in different locations. Therefore, the process of collecting persistent data from sparse data repositories is very valuable. In this study, artificial neural network systems such as ANN, has been proposed for air quality prediction based on previous studies. However, these methods were not successful in utilizing existing air quality big data on temporal and statistical data features (Arumugam et al., 2023; Saravanan et al. 2023). Additionally, the starling murmuration optimization algorithm is used to optimise the ANN parameters to achieve a lower forecasting error. The inferior performance of these existing methods points to their limitations and a more effective method to predict air quality is required. Therefore, there is an urgent need to create good air pollution predicting system and more importantly, create a system that has good accuracy as compared to the present air pollution predictor models is necessary (Sivakumar et al., 2023; Sundar et al., 2022).

3.1. Impacts of air pollution

The possessions of air pollutants on living organisms will not be restricted to the health of humans and animals but will include the entire situation. Different geographical conditions, global climatic shifts, and ecological variances have an effect on human health, the environment, and animal life. Air pollution affects life on earth in numerous ways. Air pollution has a negative impact on all important industries, including health, agriculture, and economy (Murugesan et al., 2023).

4. Proposed methodologies

This study also focuses on and promotes the development of smart city environments by obtaining influential pollutants that affect the air, thereby reducing the source of specific pollutants An ANN model was created in this study to forecast. In the proposed model, the Artificial Neural Network model is used as a forecaster engine to predict and the starling murmuration optimization (SMO) algorithm is used to optimise the Artificial Neural Network parameters to achieve a lower forecasting error (Shanmugamoorthy et al., 2023a; Shanmugamoorthy et al., 2023b). Furthermore, in this study, we used real time dataset as we have used Winsen ZPHS01B sensor module to collect the data, which is stored in cloud platform. After the collected data is used to train and test, after this process we will evaluate the results. In below Figure 1 represent that the proposed model design.

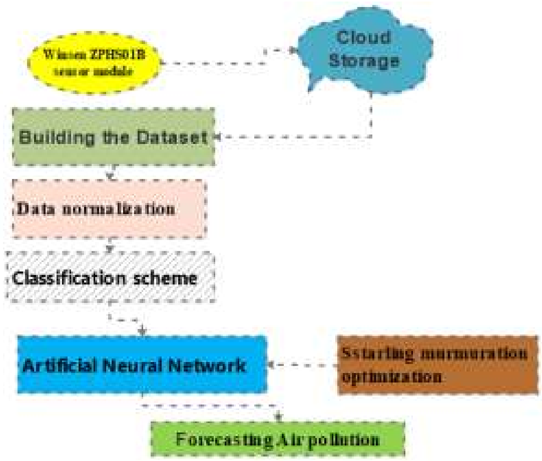
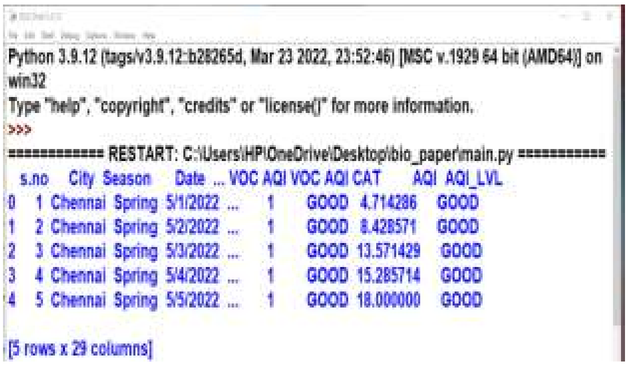


Figure 1. Proposed model architectural design

4.1. Dataset description

In this study we used real time dataset as we have used Winsen ZPHS01B sensor module to collect the data, which is stored in cloud platform. After the collected data is used to train and test, after this process we will evaluate the results (Saravanan et al., 2023). The head keyword to view my dataset description as shown Figure 2.



```
Python 3.9.12 (tags/v3.9.12:b28265d, Mar 23 2022, 23:52:46) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
============ RESTART: C:\Users\HP\OneDrive\Desktop\bio_paper\main.py ============
  s.no   City  Season   Date ... VOC AQI VOC AQI CAT        AQI  AQI_LVL
0    1 Chennai  Spring  5/1/2022 ...     1       GOOD   4.714286   GOOD
1    2 Chennai  Spring  5/2/2022 ...     1       GOOD   8.428571   GOOD
2    3 Chennai  Spring  5/3/2022 ...     1       GOOD  13.571429   GOOD
3    4 Chennai  Spring  5/4/2022 ...     1       GOOD  15.285714   GOOD
4    5 Chennai  Spring  5/5/2022 ...     1       GOOD  18.000000   GOOD

[5 rows x 29 columns]
```

Figure 2. Dataset overview

4.2. Data pre-processing phase

4.2.1. Data normalization

Zero-mean normalisation (z-score normalisation) is used to standardise the measured PM2.5 concentration data in the model provided in this study. Standard deviation standardisation is another name for zero-mean normalisation (Arumugam et al, 2023). A standard deviation of 1 can be found in the processed data, with a mean of 0. The normalisation formula is then used to get the normalised data:

$$x^{\wedge} * = (x - x^{\wedge}) / \sigma \tag{1}$$

where x is the unique data's mean and its standard deviation. When data is normalised, the standard deviation and mean may be more accurately measured by the machine learning procedures. It is also possible to derive the inverse normalisation formula, which looks like this:

$$x = \sigma x^{\wedge} * + x^{\wedge} \tag{2}$$

4.2.2. In-between of the dataset

Wavelet transform is used to break down the unique data into six smaller signal datasets. Time series prediction uses past information to anticipate future outcomes.

Both X and Y series of data may be found in each dataset. The projected data set, denoted by Y, is a series of samples. Each time window in X contains historical data samples that may be used to forecast the matching data sample in Y. X is a series of these history time windows. Time step L denotes the number of previous data samples (from X's history) that are used to make a prediction for the current data sample. In X and Y, the ith component should be:

$$\{(X\_((i)) = \{x\_i, x\_(i+1), \ldots, x\_(i+L)\} \\ @Y\_((i)) = x\_(i+L+1) \tag{3}$$

We then separate the data into a samples (950 days) were used as the training, the next 1200 samples (50 days) were used as the validation set, and the last 1200 samples (50 days) were used as the testing dataset (Arumugam et al., 2022). The ratio of the training, validation, and testing datasets is around 19:1:1.

In the prediction phase, deep learning models may learn from and make predictions on the deconstructed datasets once those datasets have been further divided and shaped.

5. Classification scheme

A reliable land identification and classification system that informs agricultural robots where and when to trigger their actuators to conduct the required action in real-time is essential for selected and land-specific treatments. For instance, weeds exhibit rapid growth and parasitically compete with legitimate crops for nutrients and space, although providing no significant value in terms of food, nutrition, or medicine. Classification and weed identification and plant seedlings are now the subjects of extensive study because inefficient procedures, such hand weeding, have resulted to large losses and growing expenses due to physical labour. As a result, this helps precision farming techniques more effectively manage weeds by adjusting the dosage of herbicides sprayed based on the density of the weed infestation (Shanmugamoorthy et al., 2022).

5.1. Artificial neural network

By mimicking logical systems such as the human brain, artificial neural networks have demonstrated their capabilities in producing powerful patterns for classification and that can handle very large intelligent tasks. Being a parallel system, it can complete the most complex operations or tasks in various fields. It can predict and detect problems without increasing the complexity of the problem.

Input: NI2: Input layer number of input nodes.

NH2: Hidden layer number of hidden nodes.

NO2: Output layer number of output nodes.

Range for weights: λi √NI for WHI & λi √NH for WOH

Calculate weights

WHI2: Weight attached with hidden layer node to input layer node (wij).

WOH2: Weight attached with output layer node to hidden layer Node (αi).

Calculate threshold values

TH2: Threshold of hidden layer (θi).

TO2: Threshold of output layer (Y)

where

$$TH2 = \sum(\sum(wijNIj\ xj)NHi)$$

$$TO2 = \sum \alpha iNH\ i$$

Set the number of epochs

Apply various training algorithms with above inputs.

Save the output

ANN keeps a hidden layer among one input layer and one output layer. The hidden layer processes the information data that is used for the next layer and finally forwards the results to the final layer, known as the output layer. To date, it is the most widely used algorithm and the most popular machine learning algorithm in artificial intelligence. It is used especially for Feed-Forward Back-Propagation (FBP), Non-Linear Autoregressive Exogenous model (NARX), with each model having different functions as compared to ANN. These effective machine learning techniques can solve complex problems in a fraction of a second, thus making human life much easier.

Throughout the procedure, the health function evaluates the solutions to ensure they are optimal for the coral reefs. The ANN approach uses a cost function to assess the health of the views for the features. Processing cost of prior features is computed using Eq. 1, where Pre i is the set of previous features as vi solution.

$$pqc_i = \sum_{u \in Pre_i \cap M} qc_u = \sum_{u \in Pre_i \cap M} \sum_{q \in Q} ef_q * C_u^q \quad (4)$$

This trifecta of expenses may be used to determine the viability of the perspectives. To utilise such functions, you don't even have to do anything. Thus, this expenditure ought to improve health outcomes. However, view maintenance cost should be incurred to retain the update if the basis relations are modified. The third price tag reflects the time and energy wasted after a feature's first nodes have been assessed several times. Therefore, we expect this expense's coefficient to be negative as well. Therefore, the viability of the perspectives may be determined using Eq. 5.

$$B_i = \left(\sum_{q \in Q} ef_q * C_i^q\right) - \left[\sum_{u \in Pre_i \cap M} \sum_{q \in Q} ef_q * C_u^q + \left(\sum_{r \in R} uf_r * C_i^r\right)\right] \quad (5)$$

The fitness of a solution is the sum of the fitness values of all the viewpoints of that solution. This means that Eq. 3 may be used to determine the ANN method's health function.

$$fM = \sum_{i \in M}\left(\sum_{q \in Q} ef_q * C_i^q\right) - \left[\sum_{u \in Pre_i \cap M} \sum_{q \varsigma Q} ef_q * C_u^q + \left(\sum_{r \varsigma R} uf_r * C_i^r\right)\right] \quad (6)$$

After the health function is established, the ANN procedure, which is derived from the coral reefs optimization technique, will be discussed in detail. The ANN procedure consists of the following steps:.

Rectified Linear Unit (ReLU) is the function of choice in settings.

$$\sigma(x) = \max(0, W^T + b) \quad (7)$$

The output loss of training samples should be measured using a loss function, and then the loss function should be improved to reduce the maximum value. As a loss function, cross-entropy is what we use.

5.2. Structure of neural network

There is three types of the layers in the neural networks namely,

5.2.1. Input layer

It is the first layer containing artificial neurons that receive input data from outside for processing and transmitting to the next layer.

5.2.2. Hidden layer

These are the middle layers lying in among the layer. Their task is to process data and convert input data to output data through network neurons. For data processing to be refined and effective, the weights are continuously updated to the output of the hidden layers.

5.2.3. Output layer

It is the last layer containing the units that have learned to respond to the data to obtain the final result. Most neural networks are fully linked and connected, thereby indicating that the hidden layer is fully linked between each neuron to the previous layer or input layer.

ANN is applied for the all-India Pollution data using back propagation learning optimized technique through scaled conjugate gradient method. while implementing sigmoid activation function is used as activation function for hidden layer as well as for output layer. Raw data are rehabilitated to [0.1, 0.8] to avoid the asymptotic consequence.

5.3. Advantages

The neural network has certain advantages in the applications over the other method used that are listed below.

1. Adaptive learning is possible with ANN because it allows you to acquire skills by studying and practising with data that is specific to your situation.
2. Self-Organization: ANN makes its own structure or shows how it uses the information it has learned.
3. Real-time process: With ANN, calculations and predictions can be done at the equivalent time using special hardware strategies that are being designed and made.
4. Pattern recognition is a powerful method that can pull information from a dataset and make it applicable to other datasets. The neural network learns to recognise patterns in the dataset.
5. Instead of being programmed, the ANN system is built by learning. Neural networks can figure out patterns in a dataset on their own, so analysts can focus on more important tasks.
6. Neural networks can adapt to a world that is always changing. Neural networks are good at adapting to information that is always changing, even if it takes them a while to learn about sudden and big changes.
7. Given the ineffectiveness of the conventional method, the neural network can instead construct a model of the data. When conventional techniques, such as inferential statistics or programming logic, fail, neural networks can model the data with ease. This is due to the fact that neural networks are capable of addressing intricate relationships.
8. Neural networks work just as well as traditional statistical models, and most people think they work better for most problems. The model made by the neural network can better show how the data is organised in less time.

6. Need for optimization

The term "optimization" refers to a collection of strategies for getting the most out of technical systems in terms of both design and implementation. An optimization algorithm is a process that runs in loops, comparing and contrasting different solutions until the best one is determined. Various existing systems have relied on tried-and-true methods for extracting texture information. When the size of repository is small, there is no necessity to consider computational complexity for retrieving the best matched images to the given query image. The computational complexity has to be taken into consideration only when the search space blows exponentially. It has become obligatory to provide an effective method to retrieve images from the enormous collections of images used in different applications. The key parameter to be calculated to find best matched images is energy fitness function. It is only the optimization algorithms that can converge the energy fitness function to a minimum value in less time.

6.1. Starling murmuration optimization (SMO) algorithm

Using the murmuration modelling given in the preceding part as a starting point, this section introduces a new bio-inspired method, the (SMO). Initially, N starlings are spread out throughout the D-dimensional search space with the help of Eq. (8), where $x\ id, x\ dU, and\ x\ dL$ are the bounds of the search space and the dth dimension of starling s i, respectively.

$$x_{id} = x_d^L + rand(0,1) \times \left(x_d^U - x_d^L\right), \quad d = 1,2,\ldots,D; i = 1,2,\ldots,N \tag{8}$$

Then, the separating search technique specified by Eq. (17) is applied to a subset of the starlings in order to investigate the issue space (22). In Algorithm, the remaining starlings are taken into account when constructing the multi-flock dynamically, with the first step being the construction of the representative set S f, which includes the starlings' fitness ratings ranked in ascending order. Definition 2 is then used to build the multi-flock. After that, we use Eq. (8) to determine whether we should use technique based on the quality of the flocks we're looking at. Each iteration concludes with an update to both the local best solution and the global best key. Algorithm 1 presents the pseudocode for the suggested SMO procedure.

starling murmuration optimization (SMO)
Input: *N (Numberofstarling), k (Numberofflocks), andMaxIt (Maximumiterations).*
Output: *The global best solution.*
Begin
Randomly distribution *N* starlings in the search space.
Set t=1.
While $t \le MaxIt$
Select a proportion of starlings by Eq. (12) and moved them using the separating strategy.
Multi − flock construction by algorithm 2.
Compute the quality of k flocks by Equation.
For $q = 1 \to k$
If $Q_q(t) \le \mu_Q(t)$ */ comparing the flock quality f_q to select movement strategy
Move the Starlings of flock f_q using the dividing strategy.
Else if
End for
Update the position of starlings and global solution.
$t = t + 1$
End while
Return the position of the best starling as a global best solution.
End

multi-flock construction.
Input: S (The set of N starlings),k (Number of flocks).
Output: Multi-flocks $f_1, \ldots, f_k$
Procedure multi-flock construction
$Sf = fitness - ascending - ordered\ representation\ of\ the\ set\ S.$
$R = the\ k\ first\ elements\ of\ S_f$, including $R_1, \ldots, R_k$
$P = the\ k\ portions\ P_1 \ldots P_k$ built from $S_f - R = \{Sf_{k+1}, Sf_{k+2}, \ldots, Sf_N\}$
For $i = 1 \to k$
$f_i = \{R_i, P_i\}$*/Constructing flock f_i using Definition 2.
End For
Return Multi-flocks $f_1, \ldots, f_k$
End Procedure

7. Experimental setup

TensorFlow and Kera's is used to implement both proposed models. The models are built using the Real time dataset. For implementing the models used 3.4 GHz Intel Core i7 processor, 32 GB-RAM and 4 GB Graphic cards. we have used Winsen ZPHS01B sensor module to collect the data, which is stored in cloud platform.

7.1. Metrics derived from the confusion matrix

Many indicators can be used to ascertain the performance of the machine learning classifiers. For the sake of brevity, the following notations were used in this study:

- TN = "True Negative"
- FP = "False Positive"
- FN = "False Negative"
- TP = "True Positive"

7.1.1. Accuracy

Accuracy is a degree of how many precise forecasts your perfect made for the whole test dataset.It can be computed using Eq. (9).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Thus, accuracy is a "good basic indicator to measure model performance". The major disadvantage of simple accuracy is that it works as strong indicator in balancing data sets and with unbalanced data sets, it become poor indicator.

7.1.2. Recall

Recall refers to the "measure for how many TPs gets predicted out of all the positives in the dataset". It is also referred to as "TP rate" or "sensitivity". It can be measured using the following Eq. (10).

$$recall = \frac{TP}{TP + FN} \tag{10}$$

The value of recall can be adjusted by adjusting various parameters or hyper-parameters of the machine learning model. Lower or higher recall rates can be achieved by adjusting these parameters. These recall rates have specific meanings for a particular model:

7.1.3. Precision

Precision is another measure that is closely linked to recall. Precision is "a way to measure how right a positive prediction is." Thus, if the predicted result turns out to be positive, precision indicates whether or not the result is positive. It can be measured using the Eq. (11).

$$precision = \frac{TP}{TP + FP} \tag{11}$$

Similar to recall, precision can be easily tuned by tuning the limitations and hyper parameters of the proposed model. While tuning, it can be observed that a higher recall leads to a lower precision and a higher precision led to a lower recall.

7.1.4. F-score

The F-score is a "method to measure the accuracy of the model based on recall and precision". The F-score can be adjusted by setting the β value. The higher the F-score, the more precise the model is. Likewise, the lower the F-score, the worse the accuracy of the model The F1 score of the various classification methods can be determined using the following equation.

$$F1score = \frac{2TP}{(2TP + FP + FN)} \tag{12}$$

In a similar vein, it shows algorithm details, development, and graphs. Figure 3 displays the results of using the plots options to build and display three separate graphs, including training performance, training state, and regression. The graph is also plotted in terms of throughout the course of 4 epochs, and it can be seen that the performance plot gains less MSE over time..

Figure 3. Graph of neural network training performance

Table 1. Performance evaluation of proposed with other models

Scheme	Classifier models	Precision	Recall	F-score	Accuracy
Machine Learning Scheme	K-Nearest Neighbours	74.00	89.77	82.85	81.27
	Support Vector Machine	84.78	85.88	83.91	84.85
Deep proposed model	Proposed GAN model	93.94	92.92	95.62	91.36

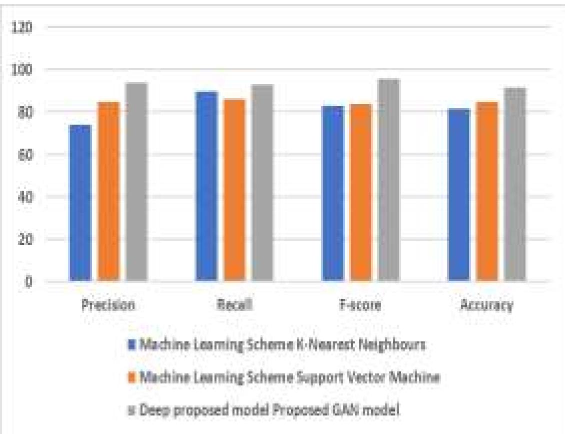


Figure 4. Performance evaluation of air pollution prediction

In the above Table 1 and Figure 4 we have compared with different classifier model to accesses the result, initially the K-Nearest Neighbour model to derive the results of accuracy is 81.27%. and Support Vector Machine network is implemented to drive the result of 94.85%. Finally, the proposed model is evaluated to measure result of 91.36% of accuracy respectively.

8. Conclusion

We have successfully implemented the simulation research of air pollution, using real-time data collected using the Winsen ZPHS01B sensor module and a cloud-based storage system. After using the collected data for training and testing, we will assess the results. To assess the effectiveness of the proposed model. The proposed technique employs an ANN to classify and forecast air pollution. In addition, the starling murmuration optimization technique is employed to optimise the parameters of the Artificial Neural Network in order to reduce predicting inaccuracy. The inadequate performance of these existing approaches demonstrates their limitations, and a more accurate method is necessary to predict air quality. In addition, the model has been evaluated using two alternative types of input parameters: type as, which includes only lagged values of the output variables. In the experimental analysis, the proposed model attained the accuracy of 91.36% respectively. When different network input variables are evaluated, the collected data suggests that the suggested model is more precise than existing benchmark models for combined forecasting.

References

Arumugam T., Kinattinkara S., Kannithottathil S., Velusamy S., Krishna M., Shanmugamoorthy M., Sivakumar V. and Boobalakrishnan K.V. (2023). Comparative assessment of groundwater quality indices of Kannur District, Kerala, India using multivariate statistical approaches and GIS, *Environmental Monitoring and Assessment*, ***195***(1), 29.

Arumugam T., Kinattinkara S., Velusamy S., Shanmugamoorthy M. and Murugan S. (2023). GIS based landslide susceptibility mapping and assessment using weighted overlay method in Wayanad: A part of Western Ghats, Kerala, *Urban Climate*, **49**, 101508.

Arumugam T., Ramachandran S., Kinattinkara S., Velusamy S., Shanmugamoorthy M. and Shanmugavadivel S. (2022). Bayesian networks and intelligence technology applied to climate change: An application of fuzzy logic based simulation in avalanche simulation risk assessment using GIS in a Western Himalayan region, *Urban Climate*, *45*, 101272.

Bai L., Wang J., Ma X. and Lu H. (2018). Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health*, **15(**4), MDPI AG, doi: 10.3390/ijerph15040780.

Cabaneros S.M., Calautit J.K.and Hughes B.R. (2019). A review of artificial neural network models for ambient air pollution prediction, *Environmental Modelling & Software,* 1, **119**, 285–304.

Chang Y.S., Chiao H.T., Abimannan S., Huang Y.P., Tsai Y.T., and Lin K.M. (2020) An LSTM-based aggregated model for air pollution forecasting, *Atmospheric Pollution Research,* **11**(8):1451–63.

Feng X., Li Q., Zhu Y., Hou J., Jin L. and Wang J. (2015) Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, *Atmospheric Environment*, **107**, 118–128, 2015, doi: 10.1016/j.atmosenv.2015.02.030.

Haviluddin F.A., Azhari M. and Ahmar A.S., (2018). Artificial neural network optimized approach for improving spatial cluster quality of land value zone, *International Journal of Engineering and Technologies*, **7(**2), 80–83, doi: 10.14419/ijet.v7i2.2.12738.

Kaimian H., Qi L., Chunlin W., Yanlin Q., Yuqin M., Gong C., Xianfeng Z. and Sonali S. (2019). Evaluation of different machine learning approaches to forecasting PM2.5 mass concentrations, Aerosol Air Quality Researchy, **19(**6), 1400–1410, doi: 10.4209/aaqr.2018.12.0450.

Kavitha S., Uma Maheswari N. and Venkatesh R. (2023). Intelligent Intrusion Detection System using Enhanced Arithmetic Optimization Algorithm with Deep Learning Model, *Technical Gazette*, **30(4),** 1217–1224. https://doi.org/10.17559/TV-20221128071759.

Mujeeb S., Alghamdi T.A., Ullah S., Fatima A., Javaid N., and Saba T. (2019). Exploiting deep learning for wind power forecasting based on big data analytics, *Applied Sciences*, , **9**(20), 4417.

Murugesan E., Shanmugamoorthy S., Veerasamy S. and Velusamy S., (2023). Groundwater hydrochemistry and its appropriateness for consumption and irrigation: Geographic and temporal variation: Integrated approach, *Urban Climate*, **49,** 101482.

Qi Y., Li Q., Karimian H., Liu D. (2019) A hybrid model for spatiotemporal forecasting of PM2. 5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment,* 10, **664**, 1–0.

Rao K.S., Devi G.L., and Ramesh N. (2019). Air quality prediction in Visakhapatnam with LSTM based recurrent neural networks, *International Journal of Intelligent System and Applications*, Feb 1, **11**(2), 18–24.

Saravanan S., Pitchaikani S., Thambiraja M., Sathiyamurthi S., Sivakumar V., Velusamy S. and Shanmugamoorthy M. (2023). Comparative assessment of groundwater vulnerability using GIS-based DRASTIC and DRASTIC-AHP for Thoothukudi District, Tamil Nadu India. *Environmental Monitoring and Assessment*, ***95***(1), 1–19.

Saravanan S., Singh L., Sathiyamurthi S., Sivakumar V., Velusamy S. and Shanmugamoorthy M. (2023). Predicting phosphorus and nitrate loads by using SWAT model in Vamanapuram River Basin, Kerala, India. *Environmental Monitoring and Assessment*, **195**(1), 186.

Shamshirband S., Hadipoor M., Baghban A., Mosavi A., Bukor J. and Várkonyi-Kóczy A.R. (2018). Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases. *Mathematics*. 1**4, 7**(10), 965.

Shanmugamoorthy M., Subbaiyan A., Elango L. and Velusamy S., (2023a). Groundwater susceptibility assessment using the GIS based DRASTIC-LU model in the Noyyal river area of South India, *Urban Climate*, **49**,101464.

Shanmugamoorthy M., Subbaiyan A., Elango L. and Velusamy S., (2023b). Groundwater Contamination Monitoring for Pollution Measurement and Transmission Applying WQI Approaches from a Region of the Erode District, Tamilnadu, India, *Journal of Water Chemistry and Technology*, ***45***(2), 181–194.

Shanmugamoorthy M., Subbaiyan A., Velusamy S. and Mani S., (2022). Review of groundwater analysis in various regions in Tamil Nadu, India. *KSCE Journal of Civil Engineering*, **26**(8), 3204–3215.

Sivakumar V., Chidambaram S.M., Velusamy S., Rathinavel R., Shanmugasundaram D.K., Sundararaj P., Shanmugamoorthy M., Thangavel R. and Balu K. (2023). An integrated approach for an impact assessment of the tank water and groundwater quality in Coimbatore region of South India: implication from anthropogenic activities, *Environmental Monitoring and Assessment*, **195**(1), 88.

Snežana S., Milesa S., Radov R., Nataša P., Nikola R., Milica S. and Dragan P. (2023). Forensic-Engineering Aspect in One Example of Long-Distance Transport of Air Pollutant, *Technical Gazette*, **30(**2), 434–440. https://doi.org/10.17559/TV-20230112000224.

Sundar M.L., Ragunath S., Hemalatha J., Vivek S., Mohanraj M., Sampathkumar V., Ansari A.M.S., Parthiban V. and Manoj S., (2022). Simulation of ground water quality for noyyal river basin of Coimbatore city, Tamilnadu using MODFLOW. *Chemosphere*, **306**, 135649.

Tao Q., Liu F., Li Y. and Sidorov D. (2019). Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU, *IEEE access*, Jun 7, **7,** 76690–8.

Thiruppathi M. and Vinoth Kumar K. (2023). Seagull Optimization-based Feature Selection with Optimal Extreme Learning Machine for Intrusion Detection in Fog Assisted WSN, *Technical Gazette*, **30(**5), 1547–1553. https://doi.org/10.17559/TV-20230130000295.

Tiwari R., Upadhyay S., Singhal P., Garg U. and Bisht S. (2019). Air pollution level prediction system, *International Journal of Innovative Technology and Exploring Engineering*.

Yan R., Liao J., Yang J., Sun W., Nong M., Li F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, 1**, 169**, 114513.

Zhang W., Wu Y. and Calautit J.K. (2022). A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. *Renewable and Sustainable Energy Reviews.* **1, 167**, 112704.

Zhao Z., Qin J., He Z., Li H., Yang Y. and Zhang R. (2020). Combining forward with recurrent neural networks for hourly air quality prediction in Northwest of China, *Environmental Science and Pollution Research*, **27(**23), 28931–28948, doi: 10.1007/s11356-020-08948-1.

Zhou Y., Chang F.J., Chang L.C., Kao I.F. and Wang Y.S. (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts, *Journal of cleaner production,* 1;**209;** 134–145.

Zhou Y., De S., Ewa G., Perera C. and Moessner K. (2021). Data-driven air quality characterisation for urban environments: A case study, *arXiv*, Accessed: Jan. **08**, [Online]. https://ieeexplore.ieee.org/abstract/document/8555540/.